

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:42:08 ; Search time 99 seconds  
(without alignments)  
1099.982 Million cell updates/sec

Title: US-10-005-549-2  
Perfect score: 2229  
Sequence: 1 MQQPQPGQQPGPGQLGG.....KSLVQIHEKNGWYTPPKEDG 422

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phage:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_rvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1448.5	65.0	378	11 Q8K228	Q8k228 mus musculus
2	1445.5	64.8	378	11 Q8BUN2	Q8bun2 mus musculus
3	1435	64.4	375	4 Q8WVN8	Q8wvn8 homo sapien
4	1308	58.7	253	11 Q9D7E1	Q9d7e1 mus musculus
5	1191	53.4	230	4 Q9BVX5	Q9bvxs5 homo sapien
6	1149.5	51.6	371	11 Q8BW45	Q8bw45 mus musculus
7	1117	50.1	217	4 Q96WV4	Q96mw4 homo sapien
8	1117	50.1	217	11 Q8K2T0	Q8k2t0 mus musculus
9	1107	49.7	255	11 Q8BVX5	Q8bvxs5 mus musculus
10	1020.5	45.8	397	5 Q9W4Z7	Q9w4z7 drosophila
11	965.5	43.3	394	5 Q46068	Q46068 drosophila
12	841	37.7	471	5 Q93571	Q93571 caenorhabdi
13	808	36.2	217	5 Q8T304	Q8t304 drosophila
14	728.5	32.7	199	4 Q96J08	Q96j08 homo sapien
15	440.5	19.8	306	5 Q9VM35	Q9vm35 drosophila
16	364	16.3	71	4 Q9UGL6	Q9ugl6 homo sapien

17	343	15.4	71	4 Q8N4G6	Q8n4g6 homo sapien
18	140	6.3	425	5 Q95039	Q95039 paramecium
19	138	6.2	829	4 Q8N8K6	Q8n8k6 homo sapien
20	136.5	6.1	199	5 Q8T0Z6	Q8t0z6 bombyx mori
21	135	6.1	199	5 P91633	P91633 drosophila
22	133	6.0	1300	12 Q36421	Q36421 alcelaphine
23	131.5	5.9	426	4 Q12937	Q12937 homo sapien
24	131.5	5.9	501	4 Q8WNB4	Q8wnb4 homo sapien
25	131.5	5.9	523	4 Q99932	Q99932 homo sapien
26	128.5	5.8	154	5 Q8T972	Q8t972 drosophila
27	128	5.7	658	11 Q9JL61	Q9jl61 mus musculus
28	128	5.7	711	4 Q00301	Q00301 homo sapien
29	127.5	5.7	207	4 Q969F4	Q969f4 homo sapien
30	127.5	5.7	207	11 Q91X63	Q91x63 mus musculus
31	127	5.7	115	4 Q9UQH5	Q9uqh5 mus musculus
32	127	5.7	342	5 Q61580	Q61580 homo sapien
33	127	5.7	525	11 Q8BIY1	Q8biy1 mus musculus
34	126	5.7	209	13 Q8AW04	Q8aw04 brachydanio
35	126	5.7	430	4 Q9H6U9	Q9h6u9 homo sapien
36	125.5	5.6	1256	11 Q99M76	Q99m76 rattus norv
37	125	5.6	1152	4 Q92603	Q92603 homo sapien
38	125	5.6	1204	4 Q00211	Q00211 homo sapien
39	125	5.6	1235	4 Q9H2G2	Q9h2g2 homo sapien
40	123.5	5.5	964	16 Q8FPA7	Q8fpa7 corynebacte
41	122.5	5.5	327	16 Q93IV2	Q93iv2 streptomyce
42	122.5	5.5	1613	11 Q8K1P7	Q8kip7 rattus norv
43	122	5.5	335	5 Q8SZS4	Q8szs4 drosophila
44	122	5.5	376	5 Q9W2X5	Q9w2x5 drosophila
45	121.5	5.5	147	10 Q9S7H6	Q9s7h6 guillardia

#### ALIGNMENTS

RESULT 1

Q8K228	PRELIMINARY;	PRT;	378 AA.
AC Q8K228			
DT 01-OCT-2002 (Tremblrel. 22, Created)			
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)			
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE Similar to NICE-5 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme			
DE -E2) (Ubiquitin-protein ligase) (Ubiquitin carrier protein).			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Strausberg R.;			
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER			
CC PROTEINS (BY SIMILARITY).			
CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +			
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLYLSINE.			
CC -!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.			
CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-			
CC THIOLESTER FORMATION (BY SIMILARITY).			
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.			
DR EMBL: BC029111; AAH29111.1; -			
DR InterPro: IPR000608; UBQ_conjugat.			
DR Pfam: PF001179; UQ_con; 1.			
DR ProDom: PD000461; UBQ_conjugat; 1.			
DR SMART: SM00212; UBQC; 1.			
KW PROSITE: PS0127; UBIQUITIN_CONJUGAT_2; 1.			
KW Ligase: ubl conjugation pathway.			
SQ SEQUENCE 378 AA; 42923 MW; DI690A9C4BC6D8DC CRC64;			

Query Match 65.0%; Score 1448.5; DB 11; Length 378;  
Best Local Similarity 74.0%; Pred. No. 1.7e-96;  
Matches 285; Conservative 33; Mismatches 52; Indels 15; Gaps 4;

QY 41 LRRELKLESIFHRGHERFRIASACLDLSCEFLLAGAGAGAGAPGPHLPGRSV-PG 99

```

Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLV-----PPPPPPGSSLSPP 54
QY     100 DPVRIHCNITESYPVPPPIWSVESDDPNLAALVRLVDIKGNLTLLQHLKRIISDLCKL 159
Db      55 PPLTLHCNITESYSSPSPIWFVSDDPNLTSLVERLED-TKNSSSLRQQLKWLICDLCLRL 113
QY     160 YNLPQHPDVMELDQPLPAEQ--CTQEDVSSDEDEEMPEDETDLDHYEMKEEPEAGKKS 217
Db      114 YNLPKHLVEMLDQPLPTGQNGTTEVTSEEEEEEAEIEDLDHYEMKEEPEINGKKS 173
QY     218 EDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNAYE 277
Db      174 EDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQSYKAGIYSVE 233
QY     278 LVNDSLYDNWVKLLKVDQDSALHNDLQILKEKEGADFIILNFSKDNFPDFPPFVRVSP 337
Db      234 LTNDSLYDNWVKLHKVDSDSLQILKEKEGIEYILLNFSKDNFPDFPPFVRVSP 293
QY     338 VLSGGVVLGGGALCMELLTKQGWSSAYSIESVIMQISATLVKGRVQFGANKSOYSLTR 397
Db      294 VLSGGVVLGGGALCMELLTKQGWSSAYSIESVIMQINATLVKGRVQFGANKNOYNLAR 353
QY     398 AQSYSKSLVQIHEKNGWYTPPKEDG 422
Db      354 AQSYSNSIVQIHEKNGWYTPPKEDG 378

RESULT 2
Q8BUN2 PRELIMINARY; .PRT; 378 AA.
AC Q8BUN2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to NICE-5 protein homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK083216; BAC38813.1; -.
SQ SEQUENCE 378 AA; 42950 MW; 4A81CA85400A1313 CRC64;

Query Match 64.8%; Score 1445.5; DB 11; Length 378;
Best Local Similarity 73.8%; Pred. No. 2.8e-96;
Matches 284; Conservative 34; Mismatches 52; Indels 15; Gaps 4;

QY     41 LRRELKLLSFHGRHERFRITASACLDLSCEFLLAGAGAGAGAAAGPHLPGRGSPG 99
Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLV-----PPPPPPGSSLSPP 54
QY     100 DPVRIHCNITESYPVPPPIWSVESDDPNLAALVRLVDIKGNLTLLQHLKRIISDLCKL 159
Db      55 PPLTLHCNITESYSSPSPIWFVSDDPNLTSLVERLED-TKNSSSLRQQLKWLICDLCLRL 113
QY     160 YNLPQHPDVMELDQPLPAEQ--CTQEDVSSDEDEEMPEDETDLDHYEMKEEPEAGKKS 217
Db      114 YNLPKHLVEMLDQPLPTGQNGTTEVTSEEEEEEAEIEDLDHYEMKEEPEINGKKS 173
QY     218 EDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNAYE 277
Db      174 EDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQSYKAGIYSVE 233
QY     278 LVNDSLYDNWVKLLKVDQDSALHNDLQILKEKEGADFIILNFSKDNFPDFPPFVRVSP 337

```

```

Db      234 LVNDSLYDNWVKLQKVDSDSLHSDLQILKEKEGIEYILLNFSKDNFPDFPPFVRVSP 293
QY     338 VLSGGVVLGGGALCMELLTKQGWSSAYSIESVIMQISATLVKGRVQFGANKSOYSLTR 397
Db      294 VLSGGVVLGGGALCMELLTKQGWSSAYSIESVIMQINATLVKGRVQFGANKNOYNLAR 353
QY     398 AQSYSKSLVQIHEKNGWYTPPKEDG 422
Db      354 AQSYSNSIVQIHEKNGWYTPPKEDG 378

RESULT 3
Q8WVN8 PRELIMINARY; .PRT; 375 AA.
AC Q8WVN8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to NICE-5 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme
DE E2) (Ubiquitin-protein ligase) (Ubiquitin carrier protein).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -> AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; BC017708; AAH17708.1; -.
DR InterPro; IPR006575; RWD.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00591; RWD; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
KW Ligase; Ub1 conjugation pathway.
SQ SEQUENCE 375 AA; 42818 MW; 7DE07315E89178A3 CRC64;

Query Match 64.4%; Score 1435; DB 4; Length 375;
Best Local Similarity 74.1%; Pred. No. 1.6e-95;
Matches 286; Conservative 33; Mismatches 47; Indels 20; Gaps 6;

QY     41 LRRELKLLSFHGRHERFRITASACLDLSCEFLLAGAGAGAGAAAGPHLPGRGSPG 99
Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLVPOQG-----SPHSPP----- 51
QY     100 DPVRIHCNITESYPVPPPIWSVESDDPNLAALVRLVDIKGNLTLLQHLKRIISDLCKL 159
Db      52 -PPLTLHCNITESYSSPSPIWFVSDDPNLTSLVERLED-TKNNNLRQQLKWLICELCSL 109
QY     160 YNLPQHPDVMELDQPLPAEQ-CTQEDVSS--EDEDEEMPEDETDLDHYEMKEEPEAGK 216
Db      110 YNLPKHLVEMLDQPLPTGQNGTTEVTSEEEEEEAEIEDLDHYEMKEEPEISGKK 169
QY     217 SEDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNAY 276
Db      170 SEDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQYKTIYSY 229
QY     277 ELVNDLSYDNWVKLLKVDQDSALHNDLQILKEKEGADFIILNFSKDNFPDFPPFVRVVS 336
Db      230 ELINDSLYDNWVKLQKVDPSLHSDLQILKEKEGIEYILLNFSKDNFPDFPPFVRVVL 289

```

```

OY 337 PVLGGYVGGGALCMELLTQGWSSAYSIESVIMQISATLVKRGVQFGANKSOYSLT 396
DB 290 PVLGGYVGGGALCMELLTQGWSSAYSIESVIMQISATLVKRGVQFGANKSOYNLA 349
OY 397 RAQSYKSLVQIHEKNGWYTPPKEDG 422
DB 350 RAQSYNSIVQIHEKNGWYTPPKEDG 375

RESULT 4
Q9D7E1 PRELIMINARY; PRT; 253 AA.
AC Q9D7E1
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Adult male tongue cDNA, RIKEN full-length enriched library,
DE clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Segabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11076861;

```

```

RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
DR EMBL; AK009324; BAB26217.2;
SQ SEQUENCE 253 AA; 28487 MW; 8FB9519FF26585CF CRC64;

Query Match 58.7%; Score 1308; DB 11; Length 253;
Best local Similarity 99.6%; Pred. No. 1.4e-86;
Matches 252; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 170 MLDQPLPAEQCTQEDYSSDEDEDEMPEDTDLHYEMKEEPAEGKSDGIGKENLAI 229
DB 1 MLDQPLPAEQCTQEDYSSDEDEDEMPEDTDLHYEMKEEPAEGKSDGIGKENLAI 60
OY 230 LEKIKKNQRODYLNAGVSSVQATDRLMKELRDIYRSQSPKGNAYVELVNDSLYDNVYK 289
DB 61 LEKIKKNQRODYLNAGVSSVQATDRLMKELRDIYRSQSPKGNAYVELVNDSLYDNVYK 120
OY 290 LLKVDQDSALHNDLQILKEGADFIILNFSFKDNFPDPFPFVVRVSPVLGGYVGGGA 349
DB 121 LLKVDQDSALHNDLQILKEGADFIILNFSFKDNFPDPFPFVVRVSPVLGGYVGGGA 180
OY 350 ICMLLTQKQWSSAYSIESVIMQISATLVKRGVQFGANKSOYSLTRAQOQSKSLVQIH 409
DB 181 ICMLLTQKQWSSAYSIESVIMQISATLVKRGVQFGANKSOYSLTRAQOQSKSLVQIH 240
OY 410 EKNGWYTPPKEDG 422
DB 241 EKNGWYTPPKEDG 253

RESULT 5
Q9BVX5 PRELIMINARY; PRT; 230 AA.
AC Q9BVX5
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical ubiquitin-protein ligase (EC 6.3.2.19) (Ubiquitin-
DE conjugating enzyme E2) (Ubiquitin carrier protein)
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -> AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; BC000848; AA000848.1;
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW Hypothetical protein; Ligase; Ub1 conjugation pathway.
FT NON_TER 1

```

Db	290	SDGYVLDGGALCMELLTNQGWSSAYSIESVILQINATLVKGRVFGVD-NHYTQVAR	348
Qy	400	QSYKSLVQIHERKNGWTPPKEDG	422
Db	349	RVYKSMVLKHEKSGWTPPKQDG	371
RESULT 7			
Q96MV4	ID	Q96MV4 PRELIMINARY; PRT; 217 AA.	
AC	Q96MV4;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Hypothetical protein FLJ18126 (SC 6.3.2.19) (Ubiquitin-conjugating enzyme E2) (Ubiquitin-protein ligase) (Ubiquitin carrier protein).		
DE	DE		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,		
RA	Hotuta T., Hirozaka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,		
RA	Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,		
RA	Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,		
RA	Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,		
RA	Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,		
RA	Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,		
RA	Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isoqai T.,		
RT	"NEDO human cDNA sequencing project.";		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER		
CC	PROTEINS (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +		
CC	DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.		
CC	-1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.		
CC	-1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-		
CC	THIOLESTER FORMATION (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.		
DR	EMBL; AK056388; BAB71173.1; ..		
DR	InterPro; IPR000608; UBQ_conjugat.		
DR	Pfam; PF001179; UBQ_con; 1.		
DR	ProDom; PD000461; UBQ_conjugat; 1.		
DR	PROSITE; PS00127; UBIQUITIN-CONJUGAT 2; 1.		
KW	Hypothetical protein; Ligase; Ub1 conjugation pathway.		
QO	SEQUENCE 217 AA; 24277 MW; 3C74AF937AE2BCCE CRC64;		
Query Match 50.1%; Score 1117; DB 4; Length 217;			
Best Local Similarity 100.0%; Pred. No. 6.8e-73;			
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	206	MKEEPAEGKSEDDGIGKENLAILEKIKKNQDYLVNGAVSGVQATRLMKELRDIYR	265
Db	1	MKEEPAEGKSEDDGIGKENLAILEKIKKNQDYLVNGAVSGVQATRLMKELRDIYR	60
Qy	266	SQSPKGGNYAVELVNDSLYDNNVKKLVQDQDSALHNDLQILKEGADFILLNFSFKDNF	325
Db	61	SQSPKGGNYAVELVNDSLYDNNVKKLVQDQDSALHNDLQILKEGADFILLNFSFKDNF	120
Qy	326	PFDPFVVRVSPVLGGYVILGGGAICMELLTKQGWSSAYSIESVIMQISATLVKGRARVQ	385
Db	121	PFDPFVVRVSPVLGGYVILGGGAICMELLTKQGWSSAYSIESVIMQISATLVKGRARVQ	180
Qy	386	FGANKSQYSLTRAQOQSKSLVQIHEKNGWTPPKEDG	422
Db	181	FGANKSQYSLTRAQOQSKSLVQIHEKNGWTPPKEDG	217
RESULT 8			
Q8K2T0	ID	Q8K2T0 PRELIMINARY; PRT; 217 AA.	



O8K2T0;  
 DT 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Similar to EG:2588.2 gene product (EC 6.3.2.19) (Ubiquitin-conjugating  
 DE enzyme E2) (Ubiquitin-protein ligase) (Ubiquitin carrier  
 DE protein).  
 GN 2310012M18RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER  
 CC PROTEINS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +  
 CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.  
 CC -!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.  
 CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-  
 CC THIOLESTER FORMATION (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.  
 DR EMBL; BC030044; AAH30044.1; -.  
 DR MGD; MGI:1917343; 2310012M18RIK.  
 DR InterPro; IPR000608; UBQ\_conjugat.  
 DR Pfam; PF00179; UQ\_con; 1.  
 DR PRODOM; PD000461; UBQ\_conjugat; 1.  
 DR SMART; SM00212; UBCC; 1.  
 DR PROSITE; PS00127; UBIQUITIN\_CONJUGAT\_2; 1.  
 DR Ligase; ubl conjugation pathway.  
 KW SEQUENCE 217 AA; 24277 MW; 3C74AF937AE2BCCE CRC64;  
 SQ

Query Match 50.1%; Score 1117; DB 11; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-73;  
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 206 MKEEPAEGKSEDDGIGKENLAILEIKKKNORODYLVNGVSGVQATDRLMKELDIYR 265  
 DB 1 MKEEPAEGKSEDDGIGKENLAILEIKKKNORODYLVNGVSGVQATDRLMKELDIYR 60  
 QY 266 SQSPKGGNYAVELVNDLSYDNVKKLVQDPSALHNDLIQLKEGADFIILNFSKDNF 325  
 DB 61 SQSPKGGNYAVELVNDLSYDNVKKLVQDPSALHNDLIQLKEGADFIILNFSKDNF 120  
 QY 326 PFDPFVRVSPVLGGVGGGATCMELLTKQGWSSAYSIESVIMQISATLVKGRARVQ 385  
 DB 121 PFDPFVRVSPVLGGVGGGATCMELLTKQGWSSAYSIESVIMQISATLVKGRARVQ 180  
 QY 386 FGANKSOYSLTRAQGSYKSLVQIHEKNGWYTPPKEDG 422  
 DB 181 FGANKSOYSLTRAQGSYKSLVQIHEKNGWYTPPKEDG 217

RESULT 9  
 Q8BVX5 PRELIMINARY; PRT; 255 AA.  
 AC Q8BVX5;  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DE Similar to NICE-5 protein homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK076148; BAC36218.1; -.  
 SQ SEQUENCE 255 AA; 28916 MW; 2F5077B3FAC2D6DF CRC64;  
 QY 170 MLDQPLPAEQ-CTQEDVSEDEDEEMPEDELDHYEMKEEPAEGKSEDDGIGKENL 227  
 DB 1 MLDQPLPTGQGTTEVTSEEEEMAEDELDHYEMKEEPEINGKKSSEDEGIEKENL 60  
 QY 228 AILEKIKKNORODYLVNGVSGVQATDRLMKELDIYRSQSPKGGNYAVELVNDLSYDN 287  
 DB 61 AILEKIKKTQDHLNGVSGVQASDRLMKELDIYRSQSRAGIYSVELINDSLYDWH 120  
 QY 288 VKLLKVDPSALHNDLIQLKEGADFIILNFSKDNFDPFVRVSPVLGGVYVGG 347  
 DB 121 VKLKVDSDSLHSDLIQLKEGIEIILLNFSKDNFDPFVRVSPVLGGVYVGG 180  
 QY 348 GAICMELLTKQGWSSAYSIESVIMQISATLVKGRARVQFANKSOYSLTRAQGSYKSLVQ 407  
 DB 181 GAIKCMELLTKQGWSSAYSIESVIMQINATLVKGRARVQFANKNOYLARAQGSYNSIVQ 240  
 QY 408 IHEKNGWYTPPKEDG 422  
 DB 241 IHEKNGWYTPPKEDG 255

RESULT 10  
 Q9W4Z7 PRELIMINARY; PRT; 397 AA.  
 ID Q9W4Z7; Q9W4Z8;  
 AC Q9W4Z7; Q9W4Z8;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 23, Last annotation update)  
 DE EG:2588.2 protein (EC 6.3.2.19).  
 GN EG:2588.2 OR CG2924.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,



Q93571		PRELIMINARY;	PRT; 471 AA.
ID	Q93571		
AC	Q93571;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	F25H2.8 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme E2)		
DE	(Ubiquitin-protein ligase) (Ubiquitin carrier protein).		
GN	F25H2.8		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;		
OC	Rhabditiidae; Pladeriniae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
FN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wilkinson J.;		
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology.";		
RL	Science 282:2012-2018(1998).		
CC	-1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER		
CC	PROTEINS (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +		
CC	DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.		
CC	-1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.		
CC	-1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-		
CC	THIOLESTER FORMATION (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.		
DR	EMBL; Z79754; CAB02096.1; -		
DR	WormPep; F25H2.8; CE09653.		
DR	InterPro; IPR002048; EF-hand.		
DR	InterPro; IPR000608; UBQ_conjugat.		
DR	Pfam; PF001179; UQ_con; 1.		
DR	ProDom; PD000461; UBQ_conjugat; 1.		
DR	SMART; SM00212; UBCc; 1.		
DR	PROSITE; PS00018; EF_HAND; 1.		
DR	PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.		
KW	Ligase; ubl conjugation pathway.		
SQ	SEQUENCE 471 AA; 54029 MW; 69EECA5B510D78 CRC64;		
Query Match 37.7%; Score 841; DB 5; Length 471;			
Best Local Similarity 44.9%; Pred No. 1.7e-52;			
Matches 180; Conservative 69; Mismatches 100; Indels 52; Gaps 10;			
QY	40 CLRR--ELKLESIFRHRHERFRIASCLDELSCFEFLAGAGGAGAAGAPGLPPRG	96	
Db	: ::::     :     :     :     :     :		
QY	3 CLRKLKDIOVLCKLPKNRNFQILSASVDELSMKFINAKNG	46	
QY	97 VPGDPVRHIHNITESYPAVPINVSIEDD-PNLAAVLERLVDKKGWTLLOHLKRISD	155	
Db	: ::::                       : : : : :		
QY	47 -----IVTANTIQENYPQPPIWFSESDPDVTGMSLQRTETEE-STNLIHQHVRLVD	100	
Db	: ::::                   : : : : :		
QY	156 LCKLYLN-----POHPDV-----EMLDQLPAEQCTCEDYSSDEDEEMP	195	
Db	: ::::     :		
QY	101 LCSFYNLQMPCLPQIAPPYRDDIDEGRSDISDTTSEPIDDDMAGGEVDDDEEDEED	160	
Db	: ::::                         : : : : :		
QY	196 EDTE-DLDHYEMKEEPAGEKKSEDGGIGKENLAILEIKKNRODYLGAVSGSVQATD	254	
Db	: ::::                         : : : : :		
QY	161 EDADGDIEIVEMAEDPT---SQHDVGSVEGLDMLDKVKINRQQHLDCKVOGSITATD	217	
Db	: ::::                         : : : : :		
QY	255 RLKKELRDIYRSOFSEKGNAYVEL-VNDSLYDNWKLLKYDQDSALHNDLQILKEKGA	313	
Db	: ::::                         : : : : :		
QY	218 RLKKELRDIHRSHEFKNGIYTFELEKEENLYQMWLKHVKYDEDSPLFDMMKLKLDHNO	277	
Db	: ::::                         : : : : :		
QY	314 FILLNFSFKONFPDPFVFRVSPVLSGGVVLGGGAICMELLTKOGWSSAYSIESVTMOI	373	
Db	: ::::                         : : : : :		
QY	278 HLFSTFTNEKFPCDPFVFRVAPHNQGVVLGGGAICMELLTKOGWSSAYSIESCILOI	337	
Db	: ::::                         : : : : :		
QY	374 SATLVKGRKVQFGA-NKSOYLSLTRAQQSKYSLVQIHENKING	413	
Db	: ::::                         : : : : :		



Search completed: July 29, 2003, 13:47:10  
Job time : 102 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 14:10:51 ; Search time 10561 seconds  
(without alignments)  
8649.310 Million cell updates/sec

Title: US-10-005-549-1

Perfect score: 2254

Sequence: 1 ctcgtctctctctacttgg.....tttgaaaaa 2254

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_htg:\*\*

3: gb\_in:\*\*

4: gb\_om:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_sts:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vl:\*\*

15: em\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_mu:\*\*

20: em\_om:\*\*

21: em\_or:\*\*

22: em\_ov:\*\*

23: em\_pat:\*\*

24: em\_ph:\*\*

25: em\_pl:\*\*

26: em\_ro:\*\*

27: em\_sts:\*\*

28: em\_un:\*\*

29: em\_vl:\*\*

30: em\_htg\_hum:\*\*

31: em\_htg\_inv:\*\*

32: em\_htg\_other:\*\*

33: em\_htg\_mus:\*\*

34: em\_htg\_pln:\*\*

35: em\_htg\_rdd:\*\*

36: em\_htg\_man:\*\*

37: em\_htg\_vrt:\*\*

38: em\_sy:\*\*

39: em\_htgo\_hum:\*\*

40: em\_htgo\_mus:\*\*

41: em\_htgo\_other:\*\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1790.4	79.4	3230	9	AY112698 Homo sapi
2	1632.4	73.3	1880	6	AX017982 Sequence
3	1632.4	73.3	1880	6	BD137288 Human nuc
4	1548.4	68.7	3094	10	AY112699 Mus muscu
5	1330	59.0	1330	6	AX017833 Sequence
6	1330	59.0	1330	6	BD137243 Human nuc
7	1160.4	51.5	1162	9	BC000848 Homo sapi
8	1115.4	49.5	2523	6	AX714199 Sequence
9	1115.4	49.5	2523	9	AK056388 Homo sapi
10	998	44.3	1173	10	BC030044 Mus muscu
11	969.4	43.0	992	9	BC009286 Homo sapi
12	880.8	39.1	885	9	HS243656 Homo sapi
13	794	35.2	808	17	AF116721 Homo sapi
14	772	34.3	772	9	BC015316 Homo sapi
15	681.8	30.2	853	10	BC051487 Mus muscu
16	567	25.2	184956	2	AC144817 Mus muscu
17	565.6	25.1	2782	10	BC029111 Mus muscu
18	557	24.7	2939	9	BC017708 Homo sapi
19	557	24.7	2969	9	HS243656 Homo sapi
20	488.4	21.7	64759	9	AL592078 Human DNA
21	478.2	21.2	239353	2	AC128996 Rattus no
22	478.2	21.2	249982	2	AC108661 Rattus no
23	461.2	20.5	2988	6	AX405769 Sequence
24	461.2	20.5	2993	9	AK000617 Homo sapi
25	442.8	19.6	62727	10	AL808107 Mouse DNA
26	432	18.2	432	6	AX410956 Sequence
27	422.6	18.7	5281	9	HSU08191 Human R kap
28	407.6	18.1	222697	2	AC094121 Rattus no
29	407.6	18.1	225763	2	AC115241 Rattus no
30	407.6	18.1	236356	2	AC128343 Rattus no
31	406	18.0	190148	2	AC102392 Mus muscu
32	388	17.2	240986	2	AC140674 Mus muscu
33	374.8	16.6	190148	2	AC102392 Mus muscu
34	371.4	16.5	585	9	AF397158 Homo sapi
35	371.4	16.5	870	6	AX384879 Sequence
36	371.4	16.5	2235	9	HUMRGEA
37	371.4	16.5	25606	2	AC010554 Homo sapi
38	371.4	16.5	42999	6	AX664499 Sequence
39	371.4	16.5	42999	9	HSU13369 Human ribos
40	371.4	16.5	43962	2	AC068192 Homo sapi
41	371.4	16.5	61490	2	AC025630 Homo sapi
42	371.4	16.5	85098	2	AL158197 Homo sapi
43	371.4	16.5	112833	2	AC068881 Homo sapi
44	371.4	16.5	136630	2	AC023572 Homo sapi
45	371.4	16.5	140653	2	AL355134 Homo sapi

# ALIGNMENTS

RESULT 1

AY112698

LOCUS

DEFINITION

AY112698

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AY112698 3230 bp mRNA linear PRI 01-JUN-2003  
Homo sapiens ubiquitin-conjugating enzyme E2Q (UBE2Q) mRNA,  
complete cds.

AY112698.1 GI:31321952

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3230)

Altman,M.E., Schulze,E., Adham,I., Koehler,M. and Engel,W.

Isolation and characterization of the human UBE2Q gene and its

Pred. No. is the number of results predicted by chance to have a



|||||  
Db 1621 TTGCTGGAACCTCAGCTGCTCCAGACTAGAGCTCTCTACCTATGATGATGATTTTA 1680  
Qy 2131 ATTTATTTCTCTATTTTCACTGACAGCTGCTTTTGGTTACAGTGTATGATGATG 2190  
Db 1681 ATTTATTTCTCTATTTTCACTGACAGCTGCTTTTGGTTACAGTGTATGATGATG 1740  
Qy 2191 TATGAAAAAATGATCTTTTGGGAAAAAATACAGTTTGTATTTGAAAA 2242  
Db 1741 TATGAAAAAATGATCTTTTGGGAAAAAATACAGTTTGTATTTGAAAA 1792  
  
RESULT 2  
AX017982  
LOCUS AX017982 1880 bp DNA linear PAT 07-SEP-2000  
DEFINITION ~~XXXXXXXXXX~~ from Patent WO9946375.  
ACCESSION AX017982  
VERSION AX017982.1 GI:10042481  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pilarsky,C.  
TITLE Human ~~XXXXXXXXXX~~ sequences from ~~XXXXXXXXXX~~ poststate tissue  
JOURNAL Patent: ~~XXXXXXXXXX~~ 217 ~~XXXXXXXXXX~~  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILARSKY CHRISTIAN (DE)  
FEATURES  
Location/Qualifiers  
1..1880  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 484 a 473 c 494 g 429 t  
ORIGIN  
|||||  
Query Match 73.3%; Score 1652.4; DB 6; Length 1880;  
Best Local Similarity 99.9%; Pred. No. 5.6e-195;  
Matches 1664; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 587 CGCGGGGGCGGGGGCGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 646  
Db 1 CGCGGGGGCGGG 60  
Qy 647 AGCTGAAGCTGCTGAGTCCATCTT-CCACCGCGGCCACGAGCGCTTCCGATTTGCCAGC 705  
Db 61 AGCTGAAGCTGCTGAGTCCATCTTGGCCACCGGGGCCACGAGCGCTTCCGATTTGCCAGC 120  
Qy 706 GCCTGCTGGAGAGCTGAGCTGCGAGTTCTGCTGGCTGGGGCGGAGGGGGGGGGGGGG 765  
Db 121 GCCTGCTGGAGAGCTGAGCTGCGAGTTCTGCTGGCTGGGGCGGAGGGGGGGGGGGGG 180  
Qy 766 GGGGCGGCGCGGGAGCCGATCTCCCGCCAGGGGGTGGTGGCTGGGGGGGGGGGGGGGG 825  
Db 181 GGGGCGGCGCGGGAGCCGATCTCCCGCCAGGGGGTGGTGGCTGGGGGGGGGGGGGGGG 240  
Qy 826 ATCCACTGCAACATCAGCGAGTCAATCCCTGCTGTCGCCCGCCATCTGGTGGTGGAGTCT 885  
Db 241 ATCCACTGCAACATCAGCGAGTCAATCCCTGCTGTCGCCCGCCATCTGGTGGTGGAGTCT 300  
Qy 886 GATGACCTAACTGGCTGCTGCTTTGGAGAGGCTGGTGGACATAAAGAAAGGAATACT 945  
Db 301 GATGACCTAACTGGCTGCTGCTTTGGAGAGGCTGGTGGACATAAAGAAAGGAATACT 360  
Qy 946 CTGCTATTGACAGCATCTGAAGAGGATCATCTCCGACCTGTGTAACTCTATAACCTCCCT 1005  
Db 361 CTGCTATTGACAGCATCTGAAGAGGATCATCTCCGACCTGTGTAACTCTATAACCTCCCT 420  
Qy 1006 CAGCATCCAGATGTGAGATGCTGGATCAACCCCTGGCAGCAGAGCAGTGCACACAGAA 1065  
|||||

Db 421 CAGCATCCAGATGTGGAGATGCTGGATCAACCCCTTCCAGCAGAGCAGTGCACACAGAA 480  
Qy 1066 GACGTGCTTTCAGAAAGATGAAGATGAGGAGATGCTTCCAGGACACAGAGACTTTAGATCAC 1125  
Db 481 GACGTGCTTTCAGAAAGATGAAGATGAGGAGATGCTTCCAGGACACAGAGACTTTAGATCAC 540  
Qy 1126 TATGAATGAAGAGGAGAGCAGCTGAGGGCAGAGAAATCTGAGATGATGCGCATTTGA 1185  
Db 541 TATGAATGAAGAGGAGAGCAGCTGAGGGCAGAGAAATCTGAGATGATGCGCATTTGA 600  
Qy 1186 AAAGAAACTTTGGCCATCTTAGAGAAATTTAAAGAAACAGAGGCGAAGATTACTTAAAT 1245  
Db 601 AAAGAAACTTTGGCCATCTTAGAGAAATTTAAAGAAACAGAGGCGAAGATTACTTAAAT 660  
Qy 1246 GGTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305  
Db 661 GGTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Qy 1306 TACCGATCAGAGATTTTCAAGGGGAAACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365  
Db 721 TACCGATCAGAGATTTTCAAGGGGAAACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
Qy 1366 TATGATTTGGAATGTCAAACTCTCAAAAGTTGACCGAGCAGCGCTTTGCAACAGATCTC 1425  
Db 781 TATGATTTGGAATGTCAAACTCTCAAAAGTTGACCGAGCAGCGCTTTGCAACAGATCTC 840  
Qy 1426 CAGATCTCTCAAGAGAAAGAGGCGGACTTCTTACTTACTTACTTACTTACTTACTTAAAGAT 1485  
Db 841 CAGATCTCTCAAGAGAAAGAGGCGGACTTCTTACTTACTTACTTACTTACTTACTTAAAGAT 900  
Qy 1486 AACTTTCCCTTTGACCCACCATTTGTCAGGGTTGTCCTCCAGCTCTCTCGAGGGTAT 1545  
Db 901 AACTTTCCCTTTGACCCACCATTTGTCAGGGTTGTCCTCCAGCTCTCTCGAGGGTAT 960  
Qy 1546 GTTCTGGCGGAGGGGGCGATCTGCTGAACTTCTCACCAGAGGGCTGAGCAGTGCC 1605  
Db 961 GTTCTGGCGGAGGGGGCGATCTGCTGAACTTCTCACCAGAGGGCTGAGCAGTGCC 1020  
Qy 1606 TACTCCATAGAGTCAGTGTATGTCAGATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1665  
Db 1021 TACTCCATAGAGTCAGTGTATGTCAGATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
Qy 1666 GTGAGTTTGGAGCCCAAAATCTCAATACAGTGTGACAGAGCAGCAGTCTCTCAAG 1725  
Db 1081 GTGAGTTTGGAGCCCAAAATCTCAATACAGTGTGACAGAGCAGCAGTCTCTCAAG 1140  
Qy 1726 TCCTTGTGTGAGATCCACGAAAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1785  
Db 1141 TCCTTGTGTGAGATCCACGAAAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
Qy 1786 CCCTTGTGTGAGATCCACGAAAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1845  
Db 1201 CCCTTGTGTGAGATCCACGAAAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
Qy 1846 TGTATTTGGATCTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905  
Db 1261 TGTATTTGGATCTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
Qy 1906 CTCTGCTTATTTGAGGACAATGATGCTTATTTAAAGCGCTTAAAGAAAAAACAACACA 1965  
Db 1321 CTCTGCTTATTTGAGGACAATGATGCTTATTTAAAGCGCTTAAAGAAAAAACAACACA 1380  
Qy 1966 GAATGTTTCAAGTACTCAAGACTGCTTACAGACCAACCAACCACTTCTGCTGGAACCT 2025  
Db 1381 GAATGTTTCAAGTACTCAAGACTGCTTACAGACCAACCAACCACTTCTGCTGGAACCT 1440  
Qy 2026 TGCTAGCAGCATCTTTATATAAGAAACTTTTCGAGCTCCTTATATTTGCTGGAACCTAG 2085  
Db 1441 TGCTAGCAGCATCTTTATATAAGAAACTTTTCGAGCTCCTTATATTTGCTGGAACCTAG 1500  
Qy 2086 CTGTGCTCCAGACTAGAGCTCTTACCTATGCTATGATGATTTTAAATTTTCTCTTA 2145  
Db 1501 CTGTGCTCCAGACTAGAGCTCTTACCTATGCTATGATGATTTTAAATTTTCTCTTA 1560







Db 1030 CGAGACTTCATCCTACTTCTCCCTTTAAAGATAAATCTCCCTTTGACCCACCGTTC 1089  
QY 1510 CTCAGGGTGTGTCTCCAGTCTCTCTCGAGGGTATGTTCTGGCGAGGGCGGCATCTGC 1569  
Db 1090 CTCAGGGTGTGTCTCCAGTCTCTCTCGAGGGTATGTTCTGGTGGAGGTGCCATCTGC 1149  
QY 1570 ATGGAACCTTCACCAAAACAGAGGCTGGAGAGTGCCTACTCCATAGATCAGTGCATG 1629  
Db 1150 ATGGAACCTTCACCAAAACAGAGGCTGGAGAGTGCCTACTCCATAGATCAGTGCATG 1209  
QY 1630 CAGATCAGTGCACACTGCTGAGGGGAAACAGCAGAGTGCATTTGGAGCCAAACAATCT 1689  
Db 1210 CAGATCAGTGCACACTGCTGAGGGGAAACAGCAGAGTGCATTTGGAGCCAAACAATCT 1269  
QY 1690 CAATACAGTCTGACAAGAGCAGCAGTCTCTACAGTCTCTGGTGCAGATCCACGAAAA 1749  
Db 1270 CAGTATAGCTGCAGAGAGCAGCAGTCTCTACAGTCTCTGGTGCAGATCCATGAAAA 1329  
QY 1750 AACGCTGGTACACACCCCAAAAGAGAGCGGTAAACCTCGAGATCACCTTCCTGCC 1809  
Db 1330 AACGCTGGTACACACCCCAAAAGAGAGTGCCTAGCCTGGAGTCTCACCTTCCT-CC 1388  
QY 1810 TCCCAGCAGCAGTGCAGTCTTACCTTTGATGCTGATTTGGATCTCAGCTGCCTC 1869  
Db 1389 TCCCAGCAGCAGTGCAGTCTTACCTTTGATGCTGATTTGGATCTCAGCTGCCTC 1448  
QY 1870 TGTGTTTCCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTCAGGACAATGAT 1929  
Db 1449 TG-GTTTCCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTCAGGACAATGAT 1506  
QY 1930 GCTATTCTAAACGCTTAAGGAAAAAACA- - - - -ACACAGAACTGTTTCAAGT 1979  
Db 1507 GCTATTCTAAACGCTTAAGGAAAAAACAACACCAACACACTGCTCAAGT 1566  
QY 1980 ACTCAAGACTGACTTACAGACCAACCAACCACTGCTGGAGACCTTCTAGCAGGCAAT 2039  
Db 1567 ACTCAAGACTGACT- - - - -ACAGACCAACCACTGCTGGAGACCTTCTAGCAGGCAAT 1621  
QY 2040 CTTATAAAGAAACTTTGAGCCTCTTATATTGCTGGAACACTCAGCTGCTCCAGACT 2099  
Db 1622 CTTAT-AAAGAACTCTCAGCCTCTTATATTGCTGGAACACTCAGCTGCTCCAGACT 1680  
QY 2100 AGAGCCTCTTACCTATGCTATGAGTATTTTAAATTTTCTTATTTTCAATGACACTG 2159  
Db 1681 AGAGCCTCTTACCTATGCTATGAGTATTTTAAATTTTCTTATTTTCTTATTTTCAATG 1740  
QY 2160 CTTTCTTTGGTTACAGTCTATGATGATGCTATCAAAAATGCTATCTTTGGGAAACA 2219  
Db 1741 CTTTCTTTGGTTTCAAGTCTATGATGATGCTATG-AAAAATGCTATCTTTGGGAAACA 1799  
QY 2220 ATTACA-GTTTGTAAATTTGAAA 2242  
Db 1800 ATTACACGTTGTTAATTTGAAGA 1823

RESULT 5  
AX017833  
LOCUS AX017833 1330 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence from Patent W09946375.  
ACCESSION AX017833  
VERSION AX017833.1 GI:10042436  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Schmitt,A., Specht,A., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pillarsky,C.  
TITLE Human nucleic acid sequences from prostate tissue  
JOURNAL Patent 6,954,953 (2005-SEP-22)  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN

(DE); PILARSKY CHRISTIAN (DE)  
Location/Qualifiers  
1. .1330  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 394 a 304 c 303 g 329 t  
ORIGIN  
Query Match 59.0%; Score 1330; DB 6; Length 1330;  
Best Local Similarity 100.0%; Pred. No. 4e-155;  
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 909 CTTGAGAGGCTGTGGACATATAAGAAAGGAATFACTCTCTATTCAGCATCTGAAGAG 968  
Db 1 CTTGAGAGGCTGTGGACATATAAGAAAGGAATFACTCTCTATTCAGCATCTGAAGAG 60  
QY 969 GATCATCTCGACCTGTGTAACTCTATAAACCCTCCCTCAGATCCAGATGTGAGATGCT 1028  
Db 61 GATCATCTCGACCTGTGTAACTCTATAAACCCTCCCTCAGATCCAGATGTGAGATGCT 120  
QY 1029 GGATCAACCTTGGCCAGCAGAGCTGCACACAGGAAGAGCTGTCTTCAGAAGATGAAGA 1088  
Db 121 GGATCAACCTTGGCCAGCAGAGCTGCACACAGGAAGAGCTGTCTTCAGAAGATGAAGA 180  
QY 1089 TGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGGAAGAGCC 1148  
Db 181 TGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGGAAGAGCC 240  
QY 1149 AGCTGAGGCAAGAAATCTGAAGATGATGGCATTTGGAAGAAAGAACTTTGGCCATCTT 1208  
Db 241 AGCTGAGGCAAGAAATCTGAAGATGATGGCATTTGGAAGAAAGAACTTTGGCCATCTT 300  
QY 1209 GAAATTAAGAAAGAACAGAGGCAAGATTAATTAATGGTGCAGTGTCTGGCTCGGTGCA 1268  
Db 301 GAAATTAAGAAAGAACAGAGGCAAGATTAATTAATGGTGCAGTGTCTGGCTCGGTGCA 360  
QY 1269 GGCCACTGACCGGCTGATGAGGAGCTCAGGGATATATACCGATACAGAGATTTCAAGG 1328  
Db 361 GGCCACTGACCGGCTGATGAGGAGCTCAGGGATATATACCGATACAGAGATTTCAAGG 420  
QY 1329 CGGAACATGCTAGCTGAACTCGTGAATGACAGTCTCTGTATGATTTGGAATGTCAACCTCT 1388  
Db 421 CGGAACATGCTAGCTGAACTCGTGAATGACAGTCTCTGTATGATTTGGAATGTCAACCTCT 480  
QY 1389 CAAAGTTGACAGCAGACAGCGCTTTGCAACAGTCTCCAGATCTCCAAAGAAAGAGAGG 1448  
Db 481 CAAAGTTGACAGCAGACAGCGCTTTGCAACAGTCTCCAGATCTCCAAAGAAAGAGAGG 540  
QY 1449 AGCGACTTCATTTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTTTGACCCACCAT 1508  
Db 541 AGCGACTTCATTTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTTTGACCCACCAT 600  
QY 1509 TGTCAAGGTTGTGTCTCCAGTCTCTGAGGGTATGTTCTGGGCGAGGGGCGCATCTG 1568  
Db 601 TGTCAAGGTTGTGTCTCCAGTCTCTGAGGGTATGTTCTGGGCGAGGGGCGCATCTG 660  
QY 1569 CATGGAATCTCTACCAAAACAGGCTGAGCAGTGCCTACTCCATAGATGTCAGTGCATCAT 1628  
Db 661 CATGGAATCTCTACCAAAACAGGCTGAGCAGTGCCTACTCCATAGATGTCAGTGCATCAT 720  
QY 1629 GCAGATCAGTCCACACTGTTGAAGGGGAAAGCAGAGTGCAGTTTGGAGCCAAACAATC 1688  
Db 721 GCAGATCAGTCCACACTGTTGAAGGGGAAAGCAGAGTGCAGTTTGGAGCCAAACAATC 780  
QY 1689 TCAATACAGTCTGACAAGAGCAGCAGTCTTACAAGTCTTGGTGCAGATGCCAGAAAA 1748  
Db 781 TCAATACAGTCTGACAAGAGCAGCAGTCTTACAAGTCTTGGTGCAGATGCCAGAAAA 840  
QY 1749 AAACGGCTGTGACACACCCCAAAAGAGAGCGCTTAACCTTGGAGTATCAACCTTCTCTCC 1808  
Db 841 AAACGGCTGTGACACACCCCAAAAGAGAGCGCTTAACCTTGGAGTATCAACCTTCTCTCC 900

```
QY 1809 CTCGCCAGGACCACTGGACCAATACCTTTGAATGCTGTATTTGGATCTCACGCTGCT 1868
Db 901 CTCGCCAGGACCACTGGACCAATACCTTTGAATGCTGTATTTGGATCTCACGCTGCT 960
QY 1869 CTGTGGTTCCTCCCTCAATTTTCTGACGCTGATAGCTGCTGCTATTTGCGAGGACAATGA 1928
Db 961 CTGTGGTTCCTCCCTCAATTTTCTGACGCTGATAGCTGCTGCTATTTGCGAGGACAATGA 1020
QY 1929 TGGCTATTCTAAACGCTAAGGAAAAAACAACACACAGCAACTGTTTCAAGTACTCAAGAC 1988
Db 1021 TGGCTATTCTAAACGCTAAGGAAAAAACAACACAGCAACTGTTTCAAGTACTCAAGAC 1080
QY 1989 TGACTTACAGACCAACCAACCACTTGTCTGGAACCTTGTCTAGCAGGCAATTTCTATAAAA 2048
Db 1081 TGACTTACAGACCAACCAACCACTTGTCTGGAACCTTGTCTAGCAGGCAATTTCTATAAAA 1140
QY 2049 GAACTTTCGAGCCTCCTTATTTGCTGGAACCTGAGCTGCTGCTCCAGCTAGAGCCCTCC 2108
Db 1141 GAACTTTCGAGCCTCCTTATTTGCTGGAACCTGAGCTGCTGCTCCAGCTAGAGCCCTCC 1200
QY 2109 TTACTATGCTATGATTTTAAATTTATTTTCTTCTTATTTCTATTTCTATGCTACACTGCTTTTGTG 2168
Db 1201 TTACTATGCTATGATTTTAAATTTATTTTCTTCTTATTTCTATTTCTATGCTACACTGCTTTTGTG 1260
QY 2169 GTTACAGTATGATGATGCTATGAAAAAATGATCTTTTGGGAAAAACAATTTACAGTT 2228
Db 1261 GTTACAGTATGATGATGCTATGAAAAAATGATCTTTTGGGAAAAACAATTTACAGTT 1320
QY 2229 TGTAAATTTG 2238
Db 1321 TGTAAATTTG 1330

RESULT 6
BD137243
LOCUS
DEFINITION Human nucleic acid sequence originating in prostatic tissue.
ACCESSION BD137243
VERSION BD137243.1 GI:23232188
KEYWORDS JP 2002505878-A/4.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1330)
Specht, T., Hinzmann, B., Schmitt, A., Pilarsky, C., Dahl, E. and
Rosenthal, A.
Human nucleic acid sequence originating in prostatic tissue
Patent: JP 2002505878-A 4 26-FEB-2002;
METAGEN GESELLSCHAFT FUER GENOM FORSCHUNG MBH
OS Homo sapiens (human)
PN JP 2002505878-A/4
PD 26-FEB-2002
PF 09-MAR-1999 JP 2000535742
PR 10-MAR-1998 DE 198 11 194.0
PI THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY,
PI EDGAR DAHL,
PI ANDRE ROSENTHAL
PC C12N15/09, A61K38/00, A61K39/395, A61K48/00, A61P13/08,
PC A61P35/00,
PC C07K14/47, C07K16/18, C12N1/19, C12N1/21, C12N5/10, C12N15/02, PC
C12P21/02.
PC C12P21/08, G01N33/53, C12N15/00, A61K37/02, C12N5/00, C12N15/00 CC
Human nucleic acid sequence originating in prostatic tissue FH
key Location/Qualifiers
FT source 1. .1330
FT /organism="Homo sapiens (human)".
FT Location/Qualifiers
1. .1330
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 394 a 304 c 303 g 329 t
```

```
ORIGIN
Query Match 59.0%; Score 1330; DB 6; Length 1330;
Best Local Similarity 100.0%; Pred. No. 4e-155;
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 909 CTTGAGAGGCTGTGGACATAAAGAAAGGAATACCTCTGCTATTTGAGCATCTGGAAGAG 968
Db 1 CTTGAGAGGCTGTGGACATAAAGAAAGGAATACCTCTGCTATTTGAGCATCTGGAAGAG 60
QY 969 GATCATCTCCGACCTGTGTAACTCTATAAACCCTCCCTCAGCATCCAGATGTGGAGATGCT 1028
Db 61 GATCATCTCCGACCTGTGTAACTCTATAAACCCTCCCTCAGCATCCAGATGTGGAGATGCT 120
QY 1029 GGATCAACCCCTGCGCAGCAGAGCAGTGCACACAGGAAGACGCTGCTTCAGAGAATGAAGA 1088
Db 121 GGATCAACCCCTGCGCAGCAGAGCAGTGCACACAGGAAGACGCTGCTTCAGAGAATGAAGA 180
QY 1089 TGAGGAGATGCTCAGGACACAGAGACTTAGATCACTATGAAATGAAGAGGAAGAGCC 1148
Db 181 TGAGGAGATGCTCAGGACACAGAGACTTAGATCACTATGAAATGAAGAGGAAGAGCC 240
QY 1149 AGCTGAGGCGCAAGAAATCTGAAGATGATGGCAATTTGAAAAAGAAACTTTGGCCATCTCTAGA 1208
Db 241 AGCTGAGGCGCAAGAAATCTGAAGATGATGGCAATTTGAAAAAGAAACTTTGGCCATCTCTAGA 300
QY 1209 GAAATTTAAAGAAACAGAGGCGAAGATTACTTAAATGGTGCAGTGTCTGGCTCGGTGCA 1268
Db 301 GAAATTTAAAGAAACAGAGGCGAAGATTACTTAAATGGTGCAGTGTCTGGCTCGGTGCA 360
QY 1269 GGCCACTGACCGCTGATGAAGAGCTCAGGATATATACCGATCACAGATTTTCAAGG 1328
Db 361 GGCCACTGACCGCTGATGAAGAGCTCAGGATATATACCGATCACAGATTTTCAAGG 420
QY 1329 CGGAAACTATGCACTGCAACTCTGTAATGACAGTCTGTATGATTGGAATGTCAAACTCCT 1388
Db 421 CGGAAACTATGCACTGCAACTCTGTAATGACAGTCTGTATGATTGGAATGTCAAACTCCT 480
QY 1389 CAAAGTTGACCAAGCAGGCGTTTGCAACAGATCTCCAGATCTCCAGATCTCCAGAAAGAGG 1448
Db 481 CAAAGTTGACCAAGCAGGCGTTTGCAACAGATCTCCAGATCTCCAGAAAGAGAGG 540
QY 1449 AGCCGACTTCACTTCTACTTAACTTTTCTTAAAGATAAATCTTCCCTTTCACCCACCATT 1508
Db 541 AGCCGACTTCACTTCTACTTAACTTTTCTTAAAGATAAATCTTCCCTTTCACCCACCATT 600
QY 1509 TGTGAGGCTGTGCTCCAGTCTCTCTGGAGGATGATGTTCTGGCGGAGGGGCGCATCTG 1568
Db 601 TGTGAGGCTGTGCTCCAGTCTCTCTGGAGGATGATGTTCTGGCGGAGGGGCGCATCTG 660
QY 1569 CATGGAACCTTCTACCAAAACAGGCGTGGAGCAGTGCCTACTTCCATAGAGTCAGTGAATCAT 1628
Db 661 CATGGAACCTTCTACCAAAACAGGCGTGGAGCAGTGCCTACTTCCATAGAGTCAGTGAATCAT 720
QY 1629 GCAGATCAGTCCCACTGTTGAGGGAAGACGAGTGCAGTTCGAGTTCGAGCAACAATC 1688
Db 721 GCAGATCAGTCCCACTGTTGAGGGAAGACGAGTGCAGTTCGAGTTCGAGCAACAATC 780
QY 1689 TCAATACAGTCTGACAAAGACACAGCAGTCTCTACAACTCTTGGTGAGATCCACGAAAA 1748
Db 781 TCAATACAGTCTGACAAAGACACAGCAGTCTCTACAACTCTTGGTGAGATCCACGAAAA 840
QY 1749 AAAGCGTGTGTACACACCCCAAAAGAGAGCGGTAACCCCTGGAGTATCACCCCTTCCTCC 1808
Db 841 AAAGCGTGTGTACACACCCCAAAAGAGAGCGGTAACCCCTGGAGTATCACCCCTTCCTCC 900
QY 1809 CTCCCCAGGACCACTGGACCAATTTACTTTGAAATGCTGTATTTGGATCTCAGCTGCCT 1868
Db 901 CTCCCCAGGACCACTGGACCAATTTACTTTTGAATGCTGTATTTGGATCTCAGCTGCCT 960
QY 1869 CTGTGGTTCCTCCCTCAATTTTCTGACGCTGATAGCTGCTGCTATTTGCGAGGACAATGA 1928
Db 961 CTGTGGTTCCTCCCTCAATTTTCTGACGCTGATAGCTGCTGCTATTTGCGAGGACAATGA 1020
```









## JOURNAL

Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Marcello Bento Soares, Ph.D.

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 66 Row: a Column: 14

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

source

Location/Qualifiers

1..1173

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clones="IMAGE:105437"

/tissue\_type="Mammary gland"

/clone\_lib="Soares\_mammary\_gland\_NbMMG"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac"

<1..697

/codon\_start=2

/product="RIKEN cDNA 2310012M18 gene"

/protein\_id="AAH30044.2"

/db\_xref="GI:28175091"

/db\_xref="LocusID:70093"

/translation="EEMPEDEDLDHYEMKEEPAEGKSEDDGIGKENLAILEKIKK

NQRQYLNGAVSQVQATDRLMKELDIYRSQFKGNYAVELVNDLYDNVKKLVK

QDSALNDLIQLEKEGADFIILNFKDFNPFDPFVRVSVPLSGVYLVAGGATC

MEILLTKGWSAYSIESVIMQISATLVKGRVQFGANKSOYSLTRAQOSYKSLVQIH

EKNQWTPPEDG"

BASE COUNT 347 a 275 c 266 g 285 t

## ORIGIN

Query Match 44.3%; Score 998; DB 10; Length 1173;

Best Local Similarity 94.18; Pred. No. 4.3e-114;

Matches 1107; Conservative 0; Mismatches 50; Indels 19; Gaps 6;

QY 1089 TGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGGAGGCC 1148

1 TGAAGAGATGCTGAGGATACAGAGACCTAGATCACTATGAATGAAGAGGAGGCC 60

QY 1149 AGCTAGGCGCAAGAAATCTGAAGATGATGGCAATTTGGAAGAAACCTTGGCCATCCTAGA 1208

61 AGCGAGGCGCAAGAAATCTGAAGATGATGGCAATTTGGAAGAAACCTTGGCCATCCTAGA 120

QY 1209 GAAATTAAGAAACACAGAGCGCAAGATTAATTAATGGTGCAGTGTCTGCTCGGTGCA 1268

121 GAAATTAAGAAACACAGAGCGCAAGATTAATTAATGGTGCAGTGTCTGCTCGGTGCA 180

QY 1269 GGCCACTGACCGGTGATGAAGAGCTCAGGATATATACCGATCACAGACTTTCAAAGG 1328

181 GGCCACTGACCGGTGATGAAGAGCTCAGGATATATACCGATCACAGACTTTCAAAGG 240

QY 1329 CGGAAATATGCACTGCAACTCGTGAATGACAGTCTGTGATTTGGAATGCAACTCCT 1388

241 CGGAAATATGCACTGCAACTCGTGAATGACAGTCTGTGATTTGGAATGCAACTCCT 300

QY 1389 CAAAGTTGACGACAGAGCGCTTGGCAACAGATCTCCAGATCTCCAGATCTCCAGAGAGG 1448

301 CAAAGTTGACGACAGATAGCCCTTTGCACATGATCTTCAGATCCTGAAGGAGAGGAGG 360

1449 AGCGGACTTCATCTCTACTTAACCTTTTCTTTAAAGATAAATTTCCCTTTGACCCACACATT 1508

361 AGCAGACTTCATCTCTACTTAACCTTTTCTTTAAAGATAAATTTCCCTTTGACCCACACATT 420

1509 TGTGAGGTTGTCTCTCCAGTCTCTCTGAGGATGTTCTTGGGGGAGGGGCGCATCTG 1568

421 CGTCAGGTTGTCTCTCCAGTCTCTCTGAGGATGTTCTTGGGGGAGGGGCGCATCTG 480

1569 CATGGAATCTCTCACCACACAGGCTGGAGCAGTGCCTACTCTCCTAGAGTCACTGATCAT 1628

481 CATGGAATCTCTCACCACACAGGCTGGAGCAGTGCCTACTCTCCTAGAGTCACTGATCAT 540

1629 GCAGATCAGTGCACACTGTGTGAAGGGGAAAGCAGCAGTGCATTTGGAGCCCAACAATC 1688

541 GCAGATCAGGCGCCAGCTGTGTGAAGGGGAAAGCAGCAGTGCATTTGGAGCCCAACAATC 600

1689 TCAATACAGTCTGACAGACACAGCAGTCTCTCAAGTCTCTTGGTGCAGATCCACGAAAA 1748

601 TCAATACAGTCTGACAGACACAGCAGTCTCTCAAGTCTCTTGGTGCAGATCCACGAAAA 660

1749 AAACGGCTGTACACACCCCAAGAGAGCGCTAACCTGGAGTATCACCTTCTCTCC 1808

661 AAACGGCTGTACACACCCCAAGAGAGCGCTAACCTGGAGTATCACCTTCTCTCC-C 719

1809 CTCCCCAGGACCACTGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCGTGCCT 1868

720 CTCCCCAGGACCACTGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCGTGCCT 779

1869 CTGTGGTTCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTGAGGAGCAATGA 1928

780 CTG-GTTTCCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTGAGGAGCAATGA 838

1929 TGGCTATTCTAAAGCGCTAAGGAAAAAACA-----ACACAGAACTGTGTTCAAG 1978

839 TGGCTATTCTAAAGCGCTAAGGAAAAAACA-----ACACAGAACTGTGTTCAAG 898

1979 TACTCAAGTCTGATACAGCAACCAACCACTTGTGGAACCTTGTAGCAGCAT 2038

899 TACTCAAGTCTGAT-----ACAGACCAACCACTTGTGGAACCTTGTAGCAGCAT 953

2039 TCTTATAAAGAAACCTTTCAGGCTCTTATATATGCTGGAACCTCAGCTGTGCTCCAGAC 2098

954 TCTTAT-AAAGAACTCTCAAGCTCTTATATGCTGGAACCTCAGCTGTGCTCCAGAC 1012

2099 TAGAGCTCTTACCTATGCTATGGATTTTAAATTTTCTTCTTATTTTCTGATGACACT 2158

1013 TAGAGCTCTTACCTATGCTATGGATTTTAAATTTTCTTCTTATTTTCTGATGACACT 1072

2159 GCTTTTGTGTTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2218

1073 GCTTTTGTGTTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131

2219 AATTACAGTTTGTAAATTTGAAAAAATAAAAAA 2254

1132 AATTACAGTTTGTAAATTTGAAAAAATAAAAAA 1167

## RESULT 11

BC009286

LOCUS

DEFINITION Homo sapiens, NICE-5 protein, clone MGC:14087 IMAGE:392747, mRNA, complete cds.

ACCESSION BC009286

VERSION BC009286.1 GI:14424527

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 992)

AUTHORS Strausberg,R.



**TITLE**  
JOURNAL

**REMARK**  
COMMENT

Direct Submission  
Submitted (12-JUN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/BTP  
cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, O.L., Masliello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
Tongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IMAGE Plate: 20 Row: b Column: 16  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6688150.

**FEATURES**  
source

1..992  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:55585"  
/db\_xref="taxon:9606"  
/clone="MGC:14087 IMAGE:3927447"  
/tissue\_type="Ovary adenocarcinoma"  
/clone\_lib="NIH\_MGC\_9"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
292..507  
/codon\_start=1  
/product="NICE-5 protein"  
/protein\_id="AAH09286.1"  
/db\_xref="GI:14424528"  
/translation="MELITKQWSSAYSIESVIMQISATLVKGRVQFGANKSOYSL  
TRAQSYKSLVQIHEKNGWYTPPKEDG"  
BASE COUNT 302 a 232 c 203 g 255 t  
ORIGIN

Query Match 43.0%; Score 969.4; DB 9; Length 992;  
Best Local Similarity 99.9%; Pred. No. 1.5e-110;  
Matches 970; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1284 GATGAGGAGCTCAGGATATATACCGATCACAGATTCAAGCGGAACTATGAGT 1343  
DB 6 GAGGAGAGGCTCAGGATATATACCGATCACAGATTCAAGCGGAACTATGAGT 65  
QY 1344 CGAACTCGTGAATGACAGTCTGTATGATTGGAATGTCAAACTCTCAAAGTTGACCAAG 1403  
DB 66 CGAACTCGTGAATGACAGTCTGTATGATTGGAATGTCAAACTCTCAAAGTTGACCAAG 125  
QY 1404 CAGGCTTTGCACACGATCTCCAGATCTCTCAAGAGAAAGAGCGGACTTCATTCT 1463  
DB 126 CAGGCTTTGCACACGATCTCCAGATCTCTCAAGAGAAAGAGCGGACTTCATTCT 185  
QY 1464 ACTTAACCTTTCCCTTAAGATAACTTCCCTTTGACCCACCATTTGTCAGGGTTGTGTC 1523  
DB 186 ACTTAACCTTTCCCTTTAAGATAACTTCCCTTTGACCCACCATTTGTCAGGGTTGTGTC 245  
QY 1524 TCCAGTCTCTCTGGAGGTATGTTCTGGCGGAGGGGCCCATCTGCATGGAATCTTCAC 1583  
DB 246 TCCAGTCTCTCTGGAGGTATGTTCTGGCGGAGGGGCCCATCTGCATGGAATCTTCAC 305

QY 1584 CAAACAGGGGTGGAGCAGTCCCTACTCCATAGAGTCAAGTATCATGAGATCATGTCGAC 1643  
DB 306 CAAACAGGGGTGGAGCAGTCCCTACTCCATAGAGTCAAGTATCATGAGATCATGTCGAC 365  
QY 1644 ACTGGTCAAGGGGAAAGCAGCAGTGGAGTGGAGCCCAAAATCTCAATACAGTCTGAC 1703  
DB 366 ACTGGTGAAGGGGAAAGCAGCAGTGGAGTGGAGCCCAAAATCTCAATACAGTCTGAC 425  
QY 1704 AAGAGCACACAGTCTCTACAAAGTCCCTGGTGGAGTCCAGAGTCCAGAGTCCAGTCC 1763  
DB 426 AAGAGCACACAGTCTCTACAAAGTCCCTGGTGGAGTCCAGAGTCCAGAGTCCAGTCC 485  
QY 1764 ACCCCCAAAAGAGAGCGGTAAACCCCTGGAGTATCACCCCTCTCCCTCCCGAGCACC 1823  
DB 486 ACCCCCAAAAGAGAGCGGTAAACCCCTGGAGTATCACCCCTCTCCCTCCCGAGCACC 545  
QY 1824 TGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCTGCTCTGTGGTTCCTCC 1883  
DB 546 TGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCTGCTCTGTGGTTCCTCC 605  
QY 1884 TCATTTTCTCTGGACGCTGATAGCTCTGCTTATTCAGGACAATGATGGCTTATCTAAACG 1943  
DB 606 TCATTTTCTCTGGACGCTGATAGCTCTGCTTATTCAGGACAATGATGGCTTATCTAAACG 665  
QY 1944 CTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGTACTCAAGTACTCAAGTACT 2003  
DB 666 CTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGTACTCAAGTACTCAAGTACT 725  
QY 2004 CCACACACCTTGCTGGACCCCTTGCTAGCAGGATCTTATTAAGAAATCTTCGAGCCT 2063  
DB 726 CCACACACCTTGCTGGACCCCTTGCTAGCAGGATCTTATTAAGAAATCTTCGAGCCT 785  
QY 2064 CCTATATTTGCTGAAACTCAGCTGTGCTCCAGACTAGAGCCCTTACCTATGCTATGG 2123  
DB 786 CCTATATTTGCTGAAACTCAGCTGTGCTCCAGACTAGAGCCCTTACCTATGCTATGG 845  
QY 2124 ATTTTAAATTTATTTCTCTTATTTCAATGATGCTGCTTTTGGTGTACAGTATGAT 2183  
DB 846 ATTTTAAATTTATTTCTCTTATTTCAATGATGCTGCTTTTGGTGTACAGTATGAT 905  
QY 2184 GGATGCTATGAAAAAATCTATCTTTGGGAAAAACATTTACCTTTGTTTAAATTTGAAAA 2243  
DB 906 GGATGCTATGAAAAAATCTATCTTTGGGAAAAACATTTACAGTTTGTAAATTTGAAAA 965  
QY 2244 AAAAAAATAA 2254  
DB 966 AAAAAAATAA 976

RESULT 12  
HSA243666  
LOCUS HSA243666 885 bp mRNA linear PRI 18-MAR-2001  
DEFINITION Homo sapiens mRNA for NICE-5 protein.  
ACCESSION AJ243666  
VERSION AJ243666.1 GI:6688150  
KEYWORDS NICE-5 gene; NICE-5 protein.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Marenholz, I., Zirma, M., Fischer, D.F., Backendorf, C., Ziegler, A. and  
Mischke, D.  
TITLE Identification of human epidermal differentiation complex  
(EDC)-encoded genes by subtractive hybridization of entire YACs to  
a gridded keratinocyte cDNA library  
JOURNAL Genome Res. 11 (3), 341-355 (2001)  
MEDLINE 21154910  
PUBMED 11230159  
REFERENCE 2 (bases 1 to 885)  
AUTHORS Marenholz, I.  
TITLE Direct Submission

JOURNAL Submitted (30-JUN-1999) Marenholz I., Universitaetsklinikum  
Charite, Humboldt-Universitaet Berlin, Institut fuer Immunogenetik,  
Spandauer Damm 130, 14050 Berlin, GERMANY

FEATURES  
source  
1. 885  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21"  
/clone="115918"  
/cell\_type="primary keratinocytes"  
/clone\_lib="KER 1/2"  
1. 885  
/gene="NICE-5"  
212. 427  
/gene="NICE-5"  
/codon\_start=1  
/product="NICE-5 protein"  
/protein\_id="CAB65097.1"  
/db\_xref="GI:6688151"  
/db\_xref="SPTREMBL:Q9UGL6"  
/translation="MELTKRGWSAYSIESVIMQISATLVKGRVQFGANKSQYSL  
TRAQSYKSLVQIHEKNGWYTPPKEDG"  
881

polya\_site  
/gene="NICE-5"  
247 a 215 c 180 g 243 t

BASE COUNT 247 a 215 c 180 g 243 t

ORIGIN

Query Match 39.1%; Score 880.8; DB 9; Length 885;  
Best Local Similarity 99.8%; Pred. No. 1.4e-99;  
Matches 882; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1360 AGCTGTATGATGGAATGTCAAACCTCCTCAAAGTTGACGAGCAGCGCTTTGCACAAC 1419  
DB 2 AGCTGTATGATGGAATGTCAAACCTCCTCAAAGTTGACGAGCAGCGCTTTGCACAAC 61

QY 1420 GATCCAGATGCTCAAAGAGAAAGAGGAGCGACTTCATCTCTACTTAACCTTTCCCTT 1479  
DB 62 GATCTCCAGATGCTCAAAGAGAAAGAGGAGCGCGACTTCATCTCTACTTAACCTTTCCCTT 121

QY 1480 AAGATAACTTTCCCTTTGACCCACCATTTGTGAGGGTTGTCTCCAGTCTCTCTGGA 1539  
DB 122 AAGATAACTTTCCCTTTGACCCACCATTTGTGAGGGTTGTCTCCAGTCTCTCTGGA 181

QY 1540 GGGTATGTTCTGGGGGAGGGGCCATCTGCATGGAATCTCACCACAGAGGGCTGGAGC 1599  
DB 182 GGGTATGTTCTGGGGGAGGGGCCATCTGCATGGAATCTCACCACAGAGGGCTGGAGC 241

QY 1600 AGTGCTACTCCATAGATGTCAGTGATCATGACAGATCAGTGCCACACTGGTGAAGGGAAA 1659  
DB 242 AGTGCTACTCCATAGATGTCAGTGATCATGACAGATCAGTGCCACACTGGTGAAGGGAAA 301

QY 1660 GCACGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAAAGACACAGCAGTCC 1719  
DB 302 GCACGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAAAGACACAGCAGTCC 361

QY 1720 TACAAGTCTTGGTGCAGATCCACGAAAGAAAGAGCGTGGTACACACCCCAAGAAAGAC 1779  
DB 362 TACAAGTCTTGGTGCAGATCCACGAAAGAAAGAGCGTGGTACACACCCCAAGAAAGAC 421

QY 1780 GGCTAACCTCGAGTATCACCTTCCCTCCCTCCAGGACACACTGGACCAATACCTTT 1839  
DB 422 GGCTAACCTCGAGTATCACCTTCCCTCCCTCCAGGACACACTGGACCAATACCTTT 481

QY 1840 GAATGCTGTATTTGGATCTCAGCTGCGCTCTGTGTTCCCTCCCTCATTTTTCCTGGAGC 1899  
DB 482 GAATGCTGTATTTGGATCTCAGCTGCGCTCTGTGTTCCCTCCCTCATTTTTCCTGGAGC 541

QY 1900 TGATAGCTCTGCTTATTCAGGACAAATGATGCTATTTCTAAACGTAAGGAAAAAACA 1959  
DB 542 TGATAGCTCTGCTTATTCAGGACAAATGATGCTATTTCTAAACGTAAGGAAAAAACA 601

QY 1960 AACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTGTCTGG 2019  
DB 602 AACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTGTCTGG 661

QY 2020 AACCTTGTCTAGCAGGCAATCTTATAAAGAAACTTTTCGAGCCTCCTTATATTTGCTGGA 2079  
DB 662 AACCTTGTCTAGCAGGCAATCTTATAAAGAAACTTTTCGAGCCTCCTTATATTTGCTGGA 721

QY 2080 ACTCAGCTGTCTCCAGACTAGAGCCTCCTTACCTATGCTATGCTATGGAATTTTATTTT 2139  
DB 722 ACTCAGCTGTCTCCAGACTAGAGCCTCCTTACCTATGCTATGCTATGGAATTTTATTTT 781

QY 2140 CTCCTATTTTCATGACACTGCTTTTGGTTACAGTGTATGATGCTATGCAAAA 2199  
DB 782 CTCCTATTTTCATGACACTGCTTTTGGTTACAGTGTATGATGCTATGCAAAA 841

QY 2200 AATGTATCTTTGGGAAACAAATTACAGTTTCTTAATTTGAAAAA 2243  
DB 842 AATGTATCTTTGGGAAACAAATTACAGTTTCTTAATTTGAAAAA 885

RESULT 13  
AF116721 standard; RNA; HUM; 808 BP.  
ID AF116721  
XX AC AF116721;  
XX SV AF116721.1  
XX 24-MAY-2000 (Rel. 63, Created)  
DT 09-MAY-2001 (Rel. 67, Last updated, Version 2)  
XX DE Homo sapiens PRO3094 mRNA, complete cds.  
XX OS Homo sapiens (human)  
XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
XX OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
XX [1]  
RN 1-808  
RP Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
RA Liu M., He F.;  
RT "Functional prediction of the coding sequences of 121 new genes deduced by  
RT analysis of cDNA clones from human fetal liver";  
RL Unpublished.  
XX [2]  
RN 1-808  
RP Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
RA Liu M., He F.;  
RT Submitted (24-DEC-1998) to the EMBL/GenBank/DBJ databases.  
RL Department of Experimental Hematology, Institute of Radiation Medicine,  
RL Beijing Taiping Road 27, Beijing 100850, P. R. China  
XX SPTREMBL; Q9UGL6; Q9UGL6.  
DR Key Location/Qualifiers  
FH source 1. 808  
FT /db\_xref="taxon:9606"  
FT /organism="Homo sapiens"  
FT /clone="FLC1398"  
FT /tissue\_type="liver"  
FT /dev\_stage="fetal"  
FT 121. .336  
FT /codon\_start=1  
FT /db\_xref="SPTREMBL:Q9UGL6"  
FT /note="predicted protein of HQ3094"  
FT /product="PRO3094"  
FT /protein\_id="AAF71141.1"  
FT FT

```
FT /translation="MELLFKQGWSSAYSIESVIMQISATLVKARVQFGANKSQYSLT
FT RAQSYKSLVQIHEKNGWYTPPKEDG"
XX
SQ Sequence 808 BP; 231 A; 196 G; 159 C; 222 T; 0 other;

Query Match 35.2%; Score 794; DB 17; Length 808;
Best Local Similarity 99.9%; Pred. No. 7.4e-89;
Matches 805; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1450 GCCGACTTCATCTTACTTAACTTTCTCTTAAAGATAACTTCCCTTTGACCCACCACTTT 1509
DB 1 GCCGACTTCATCTTACTTAACTTTCTCTTAAAGATAACTTCCCTTTGACCCACCACTTT 60
QY 1510 GTCAGGTTGTCTCCAGTCTCTCTGAGGGTATGTTCTGGCGGAGGGGCGCATCTGC 1569
DB 61 GTCAGGTTGTCTCCAGTCTCTCTGAGGGTATGTTCTGGCGGAGGGGCGCATCTGC 120
QY 1570 ATGGAATCTTCACCAACAGGGGTGGAGCGTCCCTACTCTCCATAGAGTCAAGTATCATG 1629
DB 121 ATGGAATCTTCACCAACAGGGGTGGAGCGTCCCTACTCTCCATAGAGTCAAGTATCATG 180
QY 1630 CAGATCAGTGCACACTGGTGAAGGGGAAAGACGAGTGCAGTTTGGAGCCCAACAATCT 1689
DB 181 CAGATCAGTGCACACTGGTGAAGGGGAAAGACGAGTGCAGTTTGGAGCCCAACAATCT 240
QY 1690 CAATACAGTCTGACAGACACAGCAGTCTCTCAAGTCTCTGTCAGATCCACGAAAAA 1749
DB 241 CAATACAGTCTGACAGACACAGCAGTCTCTCAAGTCTCTGTCAGATCCACGAAAAA 300
QY 1750 AACGGTGGTACACACCCCAAGAGACGGCTAACCTGGAGTATCACCTTCTCTCC 1809
DB 301 AACGGTGGTACACACCCCAAGAGACGGCTAACCTGGAGTATCACCTTCTCTCC 360
QY 1810 TCCCGAGCACCCTGGACCAATACCTTTGAATGCTGTTTGAATGCTGACGCTGCTC 1869
DB 361 TCCCGAGCACCCTGGACCAATACCTTTGAATGCTGTTTGAATGCTGACGCTGCTC 420
QY 1870 TGTGTTCCCTCCCTCTATTTTCTGGAGCTGATAGTCTGCTTATGAGGATCAAGTATGAT 1929
DB 421 TGTGTTCCCTCCCTCTATTTTCTGGAGCTGATAGTCTGCTTATGAGGATCAAGTATGAT 480
QY 1930 GGCTATCTTAACGCTAAGG-AAAAAACAACACAGAACTGTTTCAAGTACTCAAGAC 1988
DB 481 GGCTATCTTAACGCTAAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAAGAC 540
QY 1989 TGACTTACAGAACCAACCACTTGTGCTGGAACCTTGTAGCAGGATCTTTATATAA 2048
DB 541 TGACTTACAGAACCAACCACTTGTGCTGGAACCTTGTAGCAGGATCTTTATATAA 600
QY 2049 GAACTTTGAGCCTCTTATATTGCTGGAACCTCAGCTGTCTCCAGACTAGAGCCTCC 2108
DB 601 GAACTTTGAGCCTCTTATATTGCTGGAACCTCAGCTGTCTCCAGACTAGAGCCTCC 660
QY 2109 TTACTATGCTATGATTTTAAATTTATTTCTTATTTCTTATTTCTATGATGATGATTTT 2168
DB 661 TTACTATGCTATGATTTTAAATTTATTTCTTATTTCTTATTTCTATGATGATGATTTT 720
QY 2169 GTTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2228
DB 721 GTTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 2229 TGTAAATTTGAAAAAATAAAAAA 2254
DB 781 TGTAAATTTGAAAAAATAAAAAA 806

RESULT 14
BC015316
LOCUS
DEFINITION Homo sapiens, NICE-5 protein, clone MGC:21081 IMAGE:4151953, mRNA, complete cds.
ACCESSION BC015316
VERSION BC015316.1 GI:15929786
```

```
KEYWORDS MGC. Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 772)
JOURNAL Strausberg R.
Direct Submission
Submitted (01-OCT-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 20 Row: b Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8923743.
FEATURES
Location/Qualifiers
source
1..772
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:55585"
/db_xref="taxon:9606"
/clone="MGC:21081 IMAGE:4151953"
/tissue_type="Brain, glioblastoma with EGFR amplification"
/clone_lib="NCI_CGAP_Brn64"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
88..303
/codon_start=1
/product="NICE-5 protein"
/protein_id="AAH15316.1"
/db_xref="GI:15929787"
/translation="MELLFKQGWSSAYSIESVIMQISATLVKARVQFGANKSQYSL
TRAQSYKSLVQIHEKNGWYTPPKEDG"
BASE COUNT 220 a 187 c 157 g 208 t
ORIGIN

Query Match 34.3%; Score 772; DB 9; Length 772;
Best Local Similarity 100.0%; Pred. No. 3.9e-86;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1483 GATAACTTCCCTTTGACCCACCACTTGTCCAGGTTGTGTCTCCAGTCTCTCTGAGGG 1542
DB 1 GATAACTTCCCTTTGACCCACCACTTGTCCAGGTTGTGTCTCCAGTCTCTCTGAGGG 60
QY 1543 TATGTTCTGGCGGAGGGGCCATCTGCATGGAATCTCTACCAACAGGGCTGGAGCAGT 1602
DB 61 TATGTTCTGGCGGAGGGGCCATCTGCATGGAATCTCTACCAACAGGGCTGGAGCAGT 120
QY 1603 GCCTACTCCATAGAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1662
DB 121 GCCTACTCCATAGAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 1663 CGAGTGCAGTTTGGAGCCAACTCTCAATACAGTCTGACAAGACACAGCAGTCTCTAC 1722
DB 181 CGAGTGCAGTTTGGAGCCAACTCTCAATACAGTCTGACAAGACACAGCAGTCTCTAC 240
QY 1723 AAGTCTTGTGTCAGATCCACGAAAAAAGCGTGTGTCACACCCCAAGAGAGCGG 1782
DB 1723 AAGTCTTGTGTCAGATCCACGAAAAAAGCGTGTGTCACACCCCAAGAGAGCGG 1782
```

<p><b>TITLE</b></p> <p>Journal</p>	<p>Direct Submission</p> <p>Submitted (28-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA</p> <p><b>REMARK</b></p> <p><b>COMMENT</b></p>
	<p>NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a></p> <p>Contact: MGC help desk</p> <p>Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a></p> <p>Tissue Procurement: Marcello Bento Soares, Ph.D.</p> <p>CDNA Library Preparation: M. Bento Soares, University of Iowa</p> <p>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)</p> <p>DNA Sequencing by: Institute for Systems Biology</p> <p><a href="http://www.systemsbiology.org">http://www.systemsbiology.org</a></p> <p>contact: <a href="mailto:amadan@systemsbiology.org">amadan@systemsbiology.org</a></p> <p>Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting</p> <p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a></p> <p>Series: IRAK Plate: 113 Row: c Column: 21</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.</p> <p><b>FEATURES</b></p> <p><b>SOURCE</b></p>
	<pre> 1. 853 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1363390" /tissue_type="thymus gland, mouse" /clone_lib="Soares_thymus_2NDMT" /lab_host="nuclon"</pre>

```

CDS               <!-- start=1 -->
                    /codon_start=1
                    /product="Similar to RIKEN CDNA 2310012M18 gene"
                    /protein_id="AAH51487.1"
                    /db_xref="GI:30277689.1"
                    /translation="GTRQLKKEKGADTILLNFSEKDNPDPDPVRRVSPVLSSGYV
                    LGGCAICMLLTKQWSSSAISTYSIMQISATLVAKGRQVFQGANKSQYSLTRAQQSY
                    KSLVQHKGHWITPPKEDG."
BASE COUNT       242 a   212 c   180 g   219 t
ORIGIN            1

```

Query Match	30.2%;	Score 681.8;	DB 10;	Length 853;
Best Local Similarity	93.3%;	Pred. No. 5.3e-75;		
Matches 783;	Conservative 0;	Mismatches 37;	Indels 19;	Gaps 6;
QY	1426	CAGATCCTCAAGAGAAAGAGAGCGCGACTTCATTCTACTTCTACTTCTTCTTCTTAAAGAT	1485	
Db	10	CAGATCCTCAAGAGAAAGAGAGCGAGACTTCATCTTACTTCTCTCTTCTTAAAGAT	69	
QY	1486	AACTTTCCTTTGACCCACCATTTGTCAGGGTTGTGCTCTCCAGTCTCTCTCTGGAGGGTAT	1545	
Db	70	AACTTTCCTTTGACCCACCGTCTCGTAGGGTTGTCTCCAGTCTCTCTCTGGAGGGTAT	129	
QY	1546	GTCTCGGGCGAGGGGCCATCTGCATGGAACCTTCTCACCACACAGGGCTGGAGCAGTGCC	1605	
Db	130	GTCTCGGGTGGAGGTGCCATCTGCATGGAACCTTCTCACCACAGGGCTGGAGCAGTGCC	189	
QY	1606	TACTTCATAGAGTCAGTGATCATGCAGATCAGTGGCCACACTGGTCAAGGGGAAAGACAGA	1665	
Db	190	TACTTCATAGATCCGTGATCATGCAGATCAGCGCCACGCTGGTGAAGGGGAAAGACAGG	249	
QY	1666	GTGCAGTTTGGAGCCAAACAAATCTCAATACAGTCTGCACAAGAGCACACGAGTCTCTACAG	1725	
Db	250	GTGCAGTTTGGAGCCAAACAAATCTCAGTATAGCCTTGACGAGAGCACACGATCTCTACAG	309	
QY	1726	TCCTTGGTGCAGATCCAGCAAAAAAAGCGCTGGTACACACCCCAAAAAAGAGACGGCTAA	1785	
Db	310	TCCTTGGTGCAGATCCATGAAAAAAAAGCGGTGGTACACACCCCAAAAAAGAGATGGCTAG	369	

```
QY 1786 CCTGGAGTATACCCCTTCCCTCCCTCCAGGACCACTGGACCAATTACCTTTGAATGC 1845
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 370 CCTGGAGTCTCACCTTCTCT-CCTCCAGGACCACTGGACCAATTACCTTTGAATGC 428
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1846 TGTATTTGGATCTCACGCTGCCCTCTGTGGTCCCTCCCTCATTTTCCCTGGACGTGATAG 1905
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 TGTGTTTGGATCTCACGCTGCCCTGTG-GTTTCCCTCCCTCATTTTCCCTGGACGTGATAG 487
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1906 CTCTGCCCTATTGGAGGACAATGATGGCTATTCTAAAGGCTAAGGAAAAAACA----- 1960
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 CTCTGCCCTATTGGAGGACAATGATGGCTATTCTAAAGGCTAAGGAAAAAACAACCCAC 547
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1961 -----ACACAGAACTGTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCTTG 2015
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 AACACACACTGAAGTCTCAAGTACTCAAGACTGACT-----ACAGACCAACCAACCTCG 602
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2016 CTGGAACCCCTGTAGCAGCATCTTATAAAGAACTTCGAGCCTCCTTATATGCT 2075
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 603 CTGGAACCCCTGTAGCAGCATCTTAT-AAAGAACTCTCAAGCCTCCTTATATGCT 661
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2076 GGAACCTCAGCTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGATGATTTTAATT 2135
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 GGAACCTCAGCTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGATGATTTTAATT 721
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2136 TTTTCTCTTATTTCATGCTACACTGCTTTTTTTTGGTTACAGTGTATGATGATGATGATGA 2195
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 722 TTTTCTCTTATTTCATGCTACACTGCTTTTTTTTGGTTACAGTGTATGATGATGATGATG- 780
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2196 AAAAAATGATCTTTGGGAAAAACAATTACAGTTTGTAAATTGAAAAAATAAAAAA 2254
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 AAAAAATGATCTTTGGGAAAAACAATTACAGTTTGTAAATTGAAAAAATAAAAAA 839
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: August 4, 2003, 19:04:22  
Job time : 10669 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:58:41 ; Search time 740 Seconds  
(without alignments)  
8222.342 Million cell updates/sec

Title: US-10-005-549-1  
Perfect score: 2254  
Sequence: 1 ctgcgtctctctacttg.....tttgaataaaaaaa 2254

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	2254	100.0	2254	AAD39359	Human RATL1d6 (reg
2	1554.4	69.0	1666	AAH33115	Human colon cancer
3	1329	59.0	1329	AAZ33482	Human prostate can
4	1172.4	52.0	6277	AAZ6859	Human ORFX ORF2414
5	720	31.9	791	AAZ75982	Human ORFX ORF1537
6	461.2	20.5	2988	ABN59773	Novel human coding
7	432	19.2	432	ABN97105	Gene #3603 used to
8	371.4	16.5	390	AAT19939	Human gene signatu

c 10	371.4	16.5	870	24	AAD31836	Human pancreatic t
c 11	371.4	16.5	968	21	AAC59573	Human secreted pro
c 12	371.4	16.5	13076	22	AAI06214	Human reproductive
c 13	371.4	16.5	13076	22	AAI07355	Human reproductive
c 14	371.4	16.5	42999	24	ABN565032	Invertebrate forag
c 15	369.8	16.4	1869	22	AAI14006	DNA to infer human
c 16	369.8	16.4	1869	22	AAI14299	Human DNA to infer
c 17	368.2	16.3	1969	24	ABN95733	Gene #2231 used to
c 18	353.4	15.7	42999	25	ABN11086	Human ribosomal RN
c 19	333	14.8	536	22	AAH35089	Human colon cancer
c 20	332.6	14.8	362	20	AAV86232	EST clone S70. Ho
c 21	331.4	14.7	1869	22	AAI14005	DNA to infer mouse
c 22	331.4	14.7	1869	22	AAI14298	Mouse DNA to infer
c 23	331.4	14.7	17384	25	ACC44721	Cosmid pFK161 nucl
c 24	330	14.6	22118	25	ACC44629	Mouse ribosomal RN
c 25	309.4	13.7	1426	21	AAZ52472	HTRM clone 3340290
c 26	298	13.2	22118	25	ABX11085	Mouse ribosomal RN
c 27	281	12.5	309	22	AAH81601	Human differential
c 28	280.4	12.4	281	24	ABV89079	Human colon cancer
c 29	276	12.2	451	21	AAH30588	Human colon cancer
c 30	276	12.2	1444	23	ABL20409	Drosophila melanog
c 31	269.4	12.0	2179	23	ABL05399	Drosophila melanog
c 32	269.4	12.0	316	24	ABN93696	Gene #194 used to
c 33	260	11.5	316	24	ABL62516	Colon adenocarcino
c 34	250.4	11.1	980	20	AAZ42109	Human endometrium
c 35	244	10.8	278	22	AAH81515	Human differential
c 36	227.4	10.1	357	24	AAZ64725	Novel human polynu
c 37	226.4	10.0	357	24	ABL79430	Human ovarian canc
c 38	224.4	10.0	1510	24	ABO55048	Human ovarian anti
c 39	220.8	9.8	435	20	AAV86609	EST clone C626. H
c 40	220.2	9.8	454	22	AAI24217	Human breast cance
c 41	220	9.8	223	24	ABL36952	Human colon tumour
c 42	220	9.8	378	18	AAV04906	Nucleotide sequenc
c 43	213	9.4	378	25	ACC44633	Mouse ribosomal RN
c 44	210.8	9.4	378	25	ABX11090	Clone no. 161, sub
c 45	194	8.6	323	23	ABV62051	Human prostate exp
			411	25	ABX44547	Bovine EST associa

## ALIGNMENTS

### RESULT 1

AAD39359  
ID AAD39359 standard; DNA; 2254 BP.

AC AAD39359;

XX 04-OCT-2002 (first entry)

XX Human RATL1d6 (regulated in activated T-lymphocyte 1d6) gene.

XX Human; ubiquitin conjugating enzyme; UBC; RATL1d6; Immune disorder;  
regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;  
tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;  
sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma;  
multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia;  
depression; epilepsy; acquired immuno deficiency syndrome; allergy;  
AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;  
myocardial infarction; renal tubular acidosis; gonadal dysgenesis;  
dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV;  
anticonvulsant; antiinflammatory; Cushing's syndrome; cardiant;  
ophthalmological; gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 517..1785

XX /\*tag= a

XX /product= "Human RATL1d6 protein"

XX WO200236741-A2.



```
QY 1501 CCACCATTTGTCAGGGTTGTCTCCACTCCTCTCTGGAGGGTATGTTCTGGGGGAGGG 1560
Db 1501 CCACCATTTGTCAGGGTTGTCTCCACTCCTCTCTGGAGGGTATGTTCTGGGGGAGGG 1560
QY 1561 GCCATCTGATGGAACTTCTCACCACAGAGGCTGGAGCAGTGCCTACTCTCCATAGAGTCA 1620
Db 1561 GCCATCTGATGGAACTTCTCACCACAGAGGCTGGAGCAGTGCCTACTCTCCATAGAGTCA 1620
QY 1621 GTGATCATGACATCAGTSCCAGTGTGTGAAGGGGAAAGCAGCAGTGCAGTGTGGAGCC 1680
Db 1621 GTGATCATGACATCAGTSCCAGTGTGTGAAGGGGAAAGCAGCAGTGCAGTGTGGAGCC 1680
QY 1681 AACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTCTCAAGTCTTGGTGCAGATC 1740
Db 1681 AACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTCTCAAGTCTTGGTGCAGATC 1740
QY 1741 CACGAAAAAAGCGCTGTACACACCCCAAGAGACGCTACCTACCTGAGTATCACC 1800
Db 1741 CACGAAAAAAGCGCTGTACACACCCCAAGAGACGCTACCTACCTGAGTATCACC 1800
QY 1801 CTCTCTCTCCCTCCAGGACCACTGACCAATACCTTTGAATGCTGTATTTGGATCTCA 1860
Db 1801 CTCTCTCTCCCTCCAGGACCACTGACCAATACCTTTGAATGCTGTATTTGGATCTCA 1860
QY 1861 CGCTGCCCTGTGGTTCCCTCCTCATTTTTCCTGGAGCGTGATAGCTCTGCTATTCGAG 1920
Db 1861 CGCTGCCCTGTGGTTCCCTCCTCATTTTTCCTGGAGCGTGATAGCTCTGCTATTCGAG 1920
QY 1921 GACATGATGCTATTTCTAAAGCTAAGGAAAAAACAACACAGAGTCTTTCAAGTA 1980
Db 1921 GACATGATGCTATTTCTAAAGCTAAGGAAAAAACAACACAGAGTCTTTCAAGTA 1980
QY 1981 CTCAGACTGACTTACAGACCAACCAACCTTCTGCTGGAACCTTGTAGCAGCAATTC 2040
Db 1981 CTCAGACTGACTTACAGACCAACCAACCTTCTGCTGGAACCTTGTAGCAGCAATTC 2040
QY 2041 TTATAAAGAAACCTTCAGGCTCTCTATATGCTGGAACCTAGCTGTGCTCAGACTA 2100
Db 2041 TTATAAAGAAACCTTCAGGCTCTCTATATGCTGGAACCTAGCTGTGCTCAGACTA 2100
QY 2101 GAGCTCTCTACCTATGCTATGATGATTTTATTTATTTCTTATTTTCATGTCACATGC 2160
Db 2101 GAGCTCTCTACCTATGCTATGATGATTTTATTTATTTTCTTATTTTCATGTCACATGC 2160
QY 2161 TTTTCTGGTTACAGTGTATGATGATGTATGAAAAAATGTATCTTTTGGGAAAAACA 2220
Db 2161 TTTTCTGGTTACAGTGTATGATGATGTATGAAAAAATGTATCTTTTGGGAAAAACA 2220
QY 2221 TTACAGTTTGTATTTTCAAAAAAATAAAAAA 2254
Db 2221 TTACAGTTTGTATTTTCAAAAAAATAAAAAA 2254
```

## RESULT 2

```
AAH33115
ID AAH33115 standard; cDNA; 1666 BP.
XX
AC AAH33115;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:171.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
```

```
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR
DR P-PSDB; AAG73684.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 2327-2328; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 1666 BP; 447 A; 414 C; 412 G; 390 T; 3 other;
```

```
Query Match 69.0%; Score 1554.4; DB 22; Length 1666;
Best Local Similarity 99.1%; Pred. No. 9,3e-288;
Matches 1572; Conservative 1; Mismatches 13; Indels 1; Gaps 1;
QY 661 GAGTCCATCTTCCACCGCGGCGCAGAGCG-CTTCGCGCATTCGACGCCCTGCGTGACGA 719
Db 1 GAGTCCATCTTCCACCGCGGCGCAGAGCGCTTCGCGCATTCGACGCCCTGCGTGACGA 60
QY 720 GCTGAGCTGGAGTTCCTGCTGCTGGGCGCGGAGGCGCGGCGGCGCGCGCGCG 779
Db 61 GCTGAGCTGGAGTTCCTGCTGCTGGGCGCGGAGGCGCGGCGGCGCGCGCGCG 120
QY 780 ACCGATCTCCCCCGCGGCGGCGTGGTGGGATCTCTGCGGATCTCCGATCCACAT 839
Db 121 GACCGCATCTCCCCCGCGGCGGCGTGGTGGGATCTCTGCGGATCTCCGATCCACAT 180
QY 840 CACGAGTCTATACCTCTGTGCCCCCGCATCTGCTGCGTGAGTCTGATGACCCCTAACTT 899
Db 181 CACGAGTCTATACCTCTGTGCCCCCGCATCTGCTGCGTGAGTCTGATGACCCCTAACTT 240
QY 900 GGCTGCTGTCTTGAGAGGCTGGTGACATATAAGAGGAAGTACTTCTGCTATTGACGA 959
Db 241 GGCTGCTGTCTTGAGAGGCTGGTGACATATAAGAGGAAGTACTTCTGCTATTGACGA 300
QY 960 TCTCAAGAGGATCATCTCCGACCTGTGTAACTCTATAACTCTCCCTCAGCATCCAGATGT 1019
Db 301 TCTCAAGAGGATCATCTCCGACCTGTGTAACTCTATAACTCTCCCTCAGCATCCAGATGT 360
QY 1020 GGAGATCTGGATCAACCCCTTGGCAGCAGAGCAGTGCACACAGAGACGCTGTCTTCAGA 1079
Db 361 GGAGATCTGGATCAACCCCTTGGCAGCAGAGCAGTGCACACAGAGACGCTGTCTTCAGA 420
QY 1080 AGATCAAGATGAGGAGATGCTGTAGGACACAGAGACTTATGATCACTATGAAATGAAGA 1139
Db 1139 AGATCAAGATGAGGAGATGCTGTAGGACACAGAGACTTATGATCACTATGAAATGAAGA 1139
```



Db 421 AGATGAAGATGAGAGATGCTGAGGACACAGAGACTTAGATCATCTATGAATGAAGA 480  
QY 1140 GGAAGAGCCAGCTGAGGCGAAGAAATCTGAAGATGATGGCATTTGGAAGAAACTTGGC 1199  
Db 481 GGAAGAGCCAGCTGAGGCGAAGAAATCTGAAGATGATGGCATTTGGAAGAAACTTGGC 540  
QY 1200 CATCTTAGAGAAAATTAAGAAACACAGAGGCAAGATTACTTAATGGTGCAGTGTCTGG 1259  
Db 541 CATCTTAGAGAAAATTAAGAAACACAGAGGCAAGATTACTTAATGGTGCAGTGTCTGG 600  
QY 1260 CTGGTGCAGCCACTGACCGCTGATGAAGAGCTCAGGGATATATACCGATCACAGAG 1319  
Db 601 CTGGTGCAGCCACTGACCGCTGATGAAGAGCTCAGGGATATATACCGATCACAGAG 660  
QY 1320 TTTCAAGGGGAACTATGCACTGCAACTCGTGAATGACAGTCTGTATGATGGAATGT 1379  
Db 661 TTTCAAGGGGAACTATGCACTGCAACTCGTGAATGACAGTCTGTATGATGGAATGT 720  
QY 1380 CAACACTCTCAAGTGTACAGAGCAGCGCTTTGCACAGATCTCCAGATCTCAAGA 1439  
Db 721 CAACACTCTCAAGTGTACAGAGCAGCGCTTTGCACAGATCTCCAGATCTCAAGA 780  
QY 1440 GAAAGAGGAGCGACTTCACTTACTTAACTTTTCTTAAAGATAACTTCCCTTTGA 1499  
Db 781 GAAAGAGGAGCGACTTCACTTACTTAACTTTTCTTAAAGATAACTTCCCTTTGA 840  
QY 1500 CCCACCATTTGTGAGGTTGTCTCCAGTCTCTCTGGAGGTTGTCTTGGCGGAGG 1559  
Db 841 CCCACCATTTGTGAGGTTGTCTCCAGTCTCTCTGGAGGTTGTCTTGGCGGAGG 900  
QY 1560 GGCCATCTGATGAACTTCTACCAAAACAGGGCTGGAGAGTGCCTACTCCATAGATC 1619  
Db 901 GGCCATCTGATGAACTTCTACCAAAACAGGGCTGGAGAGTGCCTACTCCATAGATC 960  
QY 1620 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679  
Db 961 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
QY 1680 CAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTTCAAGTCTTGGTGCAGAT 1739  
Db 1021 CAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTTCAAGTCTTGGTGCAGAT 1080  
QY 1740 CCAGAAAGAGGCTGTGTACACACACCCCAAGAGAGCGGTAACCTGGAGTATCAC 1799  
Db 1081 CCAGAAAGAGGCTGTGTACACACACCCCAAGAGAGCGGTAACCTGGAGTATCAC 1140  
QY 1800 CTTCT 1859  
Db 1141 CTTCT 1200  
QY 1860 ACCTGCTCTGTGTTTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1919  
Db 1201 ACCTGCTCTGTGTTTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260  
QY 1920 GGCAATGATGCTTATTTCAACCTTAAGGAAAGAAACAAACACAGAACTGTTCAAGT 1979  
Db 1261 GGCAATGATGCTTATTTCAACCTTAAGGAAAGAAACAAACACAGAACTGTTCAAGT 1320  
QY 1980 ACTCAAGAGTACTTACAGACCAACCAACCTTGTCTGGAACCTTGTCTGAGCAGGATT 2039  
Db 1321 ACTCAAGAGTACTTACAGACCAACCAACCTTGTCTGGAACCTTGTCTGAGCAGGATT 1380  
QY 2040 CTTTAAAGAAACTTTGAGCCCTCTTATATATGCTGGAACCTGCTGCTCCAGACT 2099  
Db 1381 CTTTAAAGAAACTTTGAGCCCTCTTATATGCTGGAACCTGCTGCTCCAGACT 1440  
QY 2100 AGAGCCCTTACCTATGCTATGATTTTAAATTTATTTCTTATTTCTTATTTCTATGTA 2159  
Db 1441 AGAGCCCTTACCTATGCTATGATTTTAAATTTATTTCTTATTTCTTATTTCTATGTA 1500  
QY 2160 CTTTATTTGTTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2219  
Db 1501 CTTTATTTGTTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560

QY 2220 ATTACAGTTTGTAAATTTGAAAAAAA 2246  
Db 1561 ATTACAGTTTGTAAATTTGAAAAAAA 1587

# RESULT 3

AA233482  
ID AA233482 standard; cdNA; 1329 BP.

XX AA233482;

XX 08-DEC-1999 (first entry)

XX Human prostate cancer-associated EST 6.

XX Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;

XX cancer; tissue specificity; human; ss.

XX Homo sapiens.

XX DE19811194-A1.

XX 16-SEP-1999.

XX 10-MAR-1998; 98DE-1011194.

XX 10-MAR-1998; 98DE-1011194.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-519629/44.

XX P-PSDB; AAY48312.

XX New nucleic acid expressed at high level in normal prostatic tissue and

XX encoded polypeptides, used to treat cancer and screen for therapeutic

XX agents

XX Claim 1a; 82; 194pp; German.

XX This invention describes novel nucleic acid sequences (A) that are  
XX expressed at high level in normal prostatic tissue. Polypeptides (1)  
XX encoded by (A) are used: (a) for identifying agents for treatment of  
XX prostatic cancer and (b) for therapy of prostate cancer, optionally  
XX where expressed by gene therapy methods. (A) is also used to isolate  
XX full-length genes (for gene therapy) and for recombinant production of  
XX (1), which can be used to raise specific antibodies. (A) are identified  
XX by assembly of ESTs (expressed sequence tags) before these are analyzed  
XX for expression pattern (tissue specificity). This approach eliminates  
XX many of the false results, as regards tissue specificity, associated  
XX with known methods that use single (usually short) ESTs. AA233477-233540  
XX represent expressed sequence tags described in the method of the  
XX invention.

XX SQ Sequence 1329 BP; 394 A; 304 C; 302 G; 329 T; 0 other;

Query Match 59.0%; Score 1329; DB 20; Length 1329;

Best Local Similarity 100.0%; Pred. No. 1,1e-244;

Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 CTTGGAGAGCGCTGGTGACATAAAGAAAGGGAATACTCTGCTATTGTCAGCATCTGAAGAG 968

Db 1 CTTGGAGAGCGCTGGTGACATAAAGAAAGGGAATACTCTGCTATTGTCAGCATCTGAAGAG 60

QY 969 GATCATCTCCGACCTGTGTAAACTCTATAACCTCCCTCAGCATCCAGATGTCGAGTGTCT 1028

Db 61 GATCATCTCCGACCTGTGTAAACTCTATAACCTCCCTCAGCATCCAGATGTCGAGTGTCT 120

QY 1029 GGATCAACCTTGGCCAGCAGAGCAGTGCACAGAGGAGACGTCCTTTCAGAGATGAAGA 1088

Db 121 GGATCAACCTTGGCCAGCAGAGCAGTGCACAGAGGAGACGTCCTTTCAGAGATGAAGA 180



CC	proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.	xx
CC	Sequence 6277 BP; 1521 A; 1720 C; 1611 G; 1423 T; 2 other;	
CC	Query Match 52.08; Score 1172.4; DB 21; Length 6277;	
CC	Best Local Similarity 80.68; Pred. No. 1.3e-214;	
CC	Matches 1455; Conservative 0; Mismatches 296; Indels 55; Gaps 5;	
QY	447 CGCGCGCGGAGCCGCGAGCGCGGAGGCTCCCGGCCCGCGCGCGCGGAGCGGAGC 506	
DB	2041 CAGACGCCGAGGCTTCGCGCCCCCTCGCCATTTCCAGCAGGCTCGACGAGCGGAGA 1982	
QY	507 GGAGCGGAGGATGCAGCGCGCAGCGCGCAGGGCGCAGCAGCAGCGGGGCCGGGCGAGCA 566	
DB	1981 GC CGCGGCCAGCGCGCGCGCGGGGGTGC CGCGCGGTGACGCGGCTCCGGGCCCGGCT 1922	
QY	567 GCTGGGGGCCAGGGGGCGCGCGGGGGGCCAGGGGGGGGCCAGGGGGGGGCCGGGGCC 626	
DB	1921 CCCTTTCCGCGCGCGGCTCCCTTCCGCGGCCCTCCCGCGGAGATGAGGGGAGAGTGTC 1862	
QY	627 GGGGCCCTGCTGAGGCGAGAGCTGAAGCTGCTCGAGTCCATCTCCACGCGGCCACGA 686	
DB	1861 CGTGTACGGGCTCAAGCCGAGCTGAAGTTCTCTGGCGTCCATCTTCGACGAAGCACCA 1802	
QY	687 GC GCTTCCGCAATTGCCAGCCGCTTCCCTGGACGAGCTGAGCTGGGAGTTCTCGTGGCTGG 746	
DB	1801 GCGATTCCGCACTGTCAGTTGGAACTGGACGAGCTGCACTGCCAGTTCTGTCGCGCA 1742	
QY	747 GCGCGGAGGGCGGGGGCGGGGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806	
DB	1741 GCAGGGCAGCGCGCACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1704	
QY	807 GCCTGGGGATCTCTCGCGCATCACTGCAACATCAAGGAGTTCATCCCTGCTGTGCCGCC 866	
DB	1703 -----TCCACTGCACTATCAGGAACTCTATCCATCTTCTTCCACC 1664	
QY	867 CATCTGTCGGTGAGTCTGATGACCTACTCTGGTGTGTCTTGAGAGGCTGTGTGA 926	
DB	1663 GATATGTTTGTGGATTCTGAAGACCCAAATCTGACATCATGTTCTGGAAGCTCTAGAAGA 1604	
QY	927 CATAAAGAAAGGAATACTCTGCTATTCGACGATCTCAAGAGGATCATCTCCGACCTGTG 986	
DB	1603 TACTAAGAAC---AACAAATTGCTGCTGACGAATTGAAGTGTGATGTGAATCTGTG 1547	
QY	987 TAAACTCTATAACCTCCCTCAGCATCCAGATGTGGAGATCTGGATCAACCCCTTGCACGC 1046	
DB	1546 CAGTTTATATACTTCTTAAGCACCTGGATGTGAGATCCTAGATCAACCACTACCCAC 1487	
QY	1047 AG---AGCAGTGCACACAGGAAGACGCTTCTTCAGAA-----GATGAAGATGAGGAGAT 1097	
DB	1486 GGCTCAAAATGGGACACAGAAAGTGACTTCAGAAAGAGAGGAAGAAAGAGAGAT 1427	
QY	1098 GCTGAGGACACAGAAGACTTAGATCACTATGAATCAAAAGAGAGAGAGCCACTGAGGG 1157	
DB	1426 GGCTGAAGATATAGAAGACTTAGATCACTATGATGATGAAGNAAGAGGCTATTAGTGG 1367	
QY	1158 CAAGAAATCTGAAGATGATGGCATTTGGAAAGAAACTTCGCCATCTAGAGAAATTTAA 1217	
DB	1366 GAAAACTCAGAGATGAAGAAATTTGAAAGAAATTTGGCAATATTAGAGAAATTTAA 1307	
QY	1218 AAAAGAACCCAGGCAAGATTACTTAATGGTGCAGTGTCTGGCTCGGTGCAGGCCACTGA 1277	
DB	1306 AAAGACTCAAGGCAAGACCAATTTAAATGCTGCAGTGTCTGGCTCGGTGCAGGCCACTGA 1247	
QY	1278 CCGGCTGATGAAGAGCTCAGGGAATATATACCGATCATCAGAGTTTCAAGCGCGGAACCTA 1337	

1246	CGGGCTGATGAAGAGAGCTCAGGGATATATACCGATCACAGAGTTATATAAACAGGGGATTTA	1187
1338	TGCAGTCGAACCTCGTGAATGACAGCTGTGATGATGGAAATGTCAAACCTCCTCAAAGTTGA	1397
1186	TTCAAGTGAACCTCGTGAATGACAGCTGTGATGATGGAAATGTCAAACCTCCTCAAAGTTGA	1127
1398	CCAGGACAGCGCTTTGCAACAGCATCTCCAGATTCCTAAAGAGAAAAGAGAGCCGACATT	1457
1126	CCAGGACAGCGCTTTGCAACAGCATCTCCAGATTCCTAAAGAGAAAAGAGAGCCGACATT	1067
1458	CATTCTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTTTGACCCACCAATTTGTCAAGGTT	1517
1066	CATTCTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTTTGACCCACCAATTTGTCAAGGTT	1007
1518	TGTGTCCTCCAGTCTCTCTGAGGGTATGTTCTGGCGGAGGGGCCATCTGCATGGAACCT	1577
1006	TGTGTCCTCCAGTCTCTCTGAGGGTATGTTCTGGCGGAGGGGCCATCTGCATGGAACCT	947
1578	TCTCACCAACACAGGGCTGGAGAGTGCCTTACTCCATAGAGTCAAGTATCATGCAATTCAG	1537
946	TCTCACCAACACAGGGCTGGAGAGTGCCTTACTCCATAGAGTCAAGTATCATGCAATTCAG	887
1638	TGCCACACTCGTGAAGGGGAAACACAGAGTGCAGTTTGGAGCCAAACAAATCTCAATACAG	1697
886	TGCCACACTCGTGAAGGGGAAACACAGAGTGCAGTTTGGAGCCAAACAAATCTCAATACAG	827
1698	TCTGACAAGAGCACAGCAGTCCTACAAGTCCTTGGTGCAGATCCAG - AAAAAGAGCGCT	1756
826	TCTGACAAGAGCACAGCAGTCCTACAAGTCCTTGGTGCAGATCCAG - AAAAAGAGCGCT	767
1757	GGTACACACCCCCAAAGAAAGACGGCTAACCCCTGGAGTATCACCCCTTCCCTCCCCAG	1816
766	GGTACACACCCCCAAAGAAAGACGGCTAACCCCTGGAGTATCACCCCTTCCCTCCCCAG	707
1817	GCACCACTGACCAATTAACCTTTGAATGCTGATTTGGATCTCACCGTCGCTCTGTGGTT	1876
706	GCACCACTGACCAATTAACCTTTGAATGCTGATTTGGATCTCACCGTCGCTCTGTGGTT	647
1877	CCCTCCCTCATTTTTCCTGACGCTGATAGCTCTGCCTATTGCAAGCAATGATGGCTATT	1936
646	CCCTCCCTCATTTTTCCTGACGCTGATAGCTCTGCCTATTGCAAGCAATGATGGCTATT	587
1937	CTAAACGCTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGACTGACTTAC	1996
586	CTAAACGCTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGACTGACTTAC	527
1997	AGACCAACCAACCACTTGCTGGAACCTTGCTAGCAGGCAATCTTATATAAGAAACTTT	2056
526	AGACCAACCAACCACTTGCTGGAACCTTGCTAGCAGGCAATCTTATATAAGAAACTTT	467
2057	CGAGCCCTCCTTATATTGCTGGAACCTCAGCTGTGCTCCAGACTAGAGCCCTCCCTACCTAT	2116
466	CGAGCCCTCCTTATATTGCTGGAACCTCAGCTGTGCTCCAGACTAGAGCCCTCCCTACCTAT	407
2117	GCATAGGATTTTAAATTTTCTCTTATTTTCATGTACACTGCTTTTTTTGGTTACAGT	2176
406	GCATAGGATTTTAAATTTTCTCTTATTTTCATGTACACTGCTTTTTTTGGTTACAGT	347
2177	GTATGATGGATGCTATGAAAAAATGCTATCTTTGGGAAACCAATTAACGTTTGTAAATT	2236
346	GTATGATGGATGCTATGAAAAAATGCTATCTTTGGGAAACCAATTAACGTTTGTAAATT	287
2237	TGAAAA 2242	
286	TGAAGA 281	
RESULT 5		
AAC75982		
ID	AAC75982 standard; cdna; 791 bp.	
XX	AC	AAC75982;
XX	AC	

RESULT 5  
AAC75982  
ID AAC75982 standard; cDNA; 791 bp.  
XX  
AC AAC75982;  
XX



XX WO200222660-A2.  
PN 21-MAR-2002.  
XX 10-SEP-2001; 2001WO-US26015.  
XX 11-SEP-2000; 2000US-0659671.  
XX (HYSE-) HYSEQ INC.  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
DR P-PSDB; ABB97360.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
PT  
XX  
XX Claim 1; SEQ ID NO 184; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention.  
XX  
XX Sequence 2988 BP; 880 A; 587 C; 657 G; 864 T; 0 other;  
XX  
Query Match 20.5%; Score 461.2; DB 24; Length 2988;  
Best Local Similarity 77.9%; Pred. No. 1e-78;  
Matches 556; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 1077 AGAAGATGAAGATGAGGAGTGGGCAAGAACTCTGAAGATGATGGCAATGGAAAAGAACTT 1136  
DB 675 AGAGGAAGAAGAGAGAGATGCTGAAGATATAGAGAGACTTAGATCATCTAGATGAGTAA 734  
QY 1137 AGAGGAAGAGCCAGCTGAGGCAAGAACTCTGAAGATGATGGCAATGGAAAAGAACTT 1196  
DB 735 GGAAGAAGAGCTTATTAGTGGGAAAAGTCAGAGATGAAGAAATGAAAAGAAATTT 794  
QY 1197 GGCATCTCTAGAGAAATTTAAAAGAACAGAGGCAAGATTTACTTAAATGGTGCAGTGTC 1256  
DB 795 GGCATATTAGAGAAATTTAGAGACTCAAGGCAAGACCATTAAATGGTGCAGTGTC 854  
QY 1257 TGGCTCGGTGCAGGCCACTGACCGGTGATGAAGGAGCTCAGGGATATATACCGATCACA 1316  
DB 855 TGGATCAGTGCAGAGCTTCAGATAGACTTATGAAGAGCTCAGGGATATATACAGATCACA 914  
QY 1317 GAGTTTCAAGGGGAAATATGACAGTGAATCTGGAATGACAGTCTGTATGATGGAA 1376  
DB 915 GAGTTTATAACAGGGGATTTATTCAGTGAATCATATAATGACAGTTTATATGACTGGCA 974  
QY 1377 TGTCAAACTCTCAAGTTGACAGGACAGCGCTTTGCACAGATCTCCAGATCTCTCA 1436  
DB 975 TGTAACTTGCAGAGGTTGACCCCTGATAGTCTTTGCAGAGTATCTTCAGATCTTAA 1034  
QY 1437 AGAAGAAGAGGAGCGGAGCTTCACTTCACTTAACTTTTCTTAAAGATAACTTCCCTT 1496  
DB 1035 AGAAGAAGAGCATAGATATATTGCTTAACTTCTTTTAAAGATAACTTCCAT 1094  
QY 1497 TGACCCACCATTTGTCAGGGTGTGTCCTCCAGTCTCTCTCGAGGGTATGTTCTGGCGG 1556  
DB 1095 TGATCTCCATTTGTTGAGTGTGTACCTGTCTCTCAGGAGGGTATGTTATGGGTGG 1154  
QY 1557 AGGGGCACTCTGATGGAATCTCTACCAAAACAGGGCTGGAGCTGCCTACTCCATAGA 1616

DB 1155 AGGAGCATTTATGTTGGAACTTCTCAAAAACAGGGCTGGAGCAGTGCCTACTCAATAGA 1214  
QY 1617 GTCAGTATGATGACAGATCAGTGCACACACTGGTGAAGGGGAAAGCAGCAGTGCAGTTGG 1676  
DB 1215 ATCGTCACTATGCAATTAATGCCACCTTAGTCAAGGCAAGGCAAGTGCAGTTGG 1274  
QY 1677 AGCCAAACAATCTCAATACAGTCTGACAAGAGCAGCAGTCTTACAGTCTTGGTGA 1736  
DB 1275 AGCAATTAAGAATCAATATAATAGCAAGAGCCCAACAATCTATATTCATTTGACA 1334  
QY 1737 GATCCACCAAAAACAGGCTGTACACACCCCCCAAAAAGACGCTAACCCCTG 1790  
DB 1335 GATACATGAGAAAATGCTGTGTACACCCCTCCAAAGGAAGATGCTAAATATG 1388  
RESULT 7  
ABN97105/c  
ID ABN97105 standard; DNA; 432 BP.  
XX  
AC ABN97105;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #3603 used to diagnose liver cancer.  
XX  
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumor; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX Homo sapiens.  
XX  
PN WO200229103-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 02-OCT-2001; 2001WO-US30589.  
XX  
PR 02-OCT-2000; 2000US-237054P.  
XX  
PA (GENE-) GENE LOGIC INC.  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX  
DR WPI; 2002-426119/45.  
XX  
PT Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample  
XX  
PS Claim 1; SEQ ID NO 3603; 298pp; English.  
XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 432 BP; 138 A; 76 C; 93 G; 125 T; 0 other;  
Query Match 19.2%; Score 432; DB 24; Length 432;  
Best Local Similarity 100.0%; Pred. No. 2.5e-73;

Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 GGCACCACTGGACCAATACCTTTGAATGCTGTATTGGAATCAGCTGCTCTGTGGT 1875  
 Db |||||||  
 QY 432 GGCACCACTGGACCAATACCTTTGAATGCTGTATTGGAATCAGCTGCTCTGTGGT 373  
 Db |||||||  
 QY 1876 TCCCTCCCTCATTTTCTGGACGTGATAGCTCTGCTATTGCGAGCAATGATGCTAT 1935  
 Db |||||||  
 QY 372 TCCCTCCCTCATTTTCTGGACGTGATAGCTCTGCTATTGCGAGCAATGATGCTAT 313  
 Db |||||||  
 QY 1936 TCTAACGCTAAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAAGACTGACTTA 1995  
 Db |||||||  
 QY 312 TCTAACGCTAAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAAGACTGACTTA 253  
 Db |||||||  
 QY 1996 CAGACCAACCAACCACTGCTGGAAACCTTGTGAGAGGCAATCTTATAAAGAACTT 2055  
 Db |||||||  
 QY 252 CAGACCAACCAACCACTGCTGGAAACCTTGTGAGAGGCAATCTTATAAAGAACTT 193  
 Db |||||||  
 QY 2056 TCGAGCCTCCTTATATTGCTGAAACTCAGCTGTGCTCCAGACTAGAGCCTCCTTACCTA 2115  
 Db |||||||  
 QY 192 TCGAGCCTCCTTATATTGCTGAAACTCAGCTGTGCTCCAGACTAGAGCCTCCTTACCTA 133  
 Db |||||||  
 QY 2116 TGCATGAGATTTTAAATTTATTTCTTATTTCATGATGACTGCTTTTTTGGTTACAG 2175  
 Db |||||||  
 QY 132 TGCATGAGATTTTAAATTTATTTCTTATTTCATGATGACTGCTTTTTTGGTTACAG 73  
 Db |||||||  
 QY 2176 TGTATGATGGATGCTATGAAAAAATGATATCTTTGGGAAAAACAATTTGTTTAA 2235  
 Db |||||||  
 QY 72 TGTATGATGGATGCTATGAAAAAATGATATCTTTGGGAAAAACAATTTGTTTAA 13  
 Db |||||||  
 QY 2236 TTGAAAAAATA 2247  
 Db |||||||  
 QY 12 TTGAAAAAATA 1

## RESULT 8

AAT19939  
 ID AAT19939 standard; cDNA to mRNA; 390 BP.

XX AAT19939;  
 AC AAT19939;

XX 17-JUL-1996 (first entry)  
 DT Human gene signature HUMGS01071.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

OS WO9514772-A1.

PN 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

PR (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

XX Claim 1; Page 515; 2245pp; Japanese.

XX

CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19001-196837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.

XX  
 SQ Sequence 390 BP; 109 A; 86 C; 67 G; 127 T; 1 other;

Query Match 16 58; Score 371.4; DB 16; Length 390;  
 Best Local Similarity 96.98; Pred. No. 9.6e-62;  
 Matches 378; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1854 GATCTCAGCTGCTCTGTGCTCCCTCCCTCATTTTCTCGAGCTGATAGCTCTGCT 1913  
 Db 1 GATCTCAGCTGCTCTGTGCTCCCTCCCTCATTTTCTCGAGCTGATAGCTCTGCT 60

QY 1914 ATTGAGAGCAATGATGGCTATTCTAAACGCTAAGGAAAAAACAACACAGAACTGTT 1973  
 Db 61 ATTGAGAGCAATGATGGCTATTCTAAACGCTAAGGAAAAAACAACACAGAACTGTT 120

QY 1974 TCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTGCTGGAACCTTCTAGCA 2033  
 Db 121 TCAAGTACTCAACACTGACTTACAGACCAACCAACCACTGCTGGAACCTTCTAGCA 180

QY 2034 GGCATTTCTATAAAGAAACTTTCGAGCTCTTATATTTGCTGGAACCTGCTGCTC 2093  
 Db 181 GGCATTTCTATAAAGAAACTTTCGAGCTCTTATATTTGCTGGAACCTGCTGCTC 240

QY 2094 CAGACTAGAGCTCTTACCTATGCTATGGATTTTAAATTTTCTTATTTTCATGT 2153  
 Db 241 CAGACTAGAGCTCTTACCTATGCTATGGATTTTAAATTTTCTTATTTTCATGT 300

QY 2154 ACACCTGCTTTTTTGGTTACAGTCTATGATGATGATGATGATGATGATGATGATG 2213  
 Db 301 ACACCTGCTTTTTTGGTTACAGTCTATGATGATGATGATGATGATGATGATGATG 360

QY 2214 AAAACAATTACAGTTTGTATTGTAATTTGAAAAA 2243  
 Db 361 GAACCAATTACAGTTTGTATTGTAATTTGAAAAA 390

## RESULT 9

AAD31836  
 ID AAD31836 standard; cDNA; 870 BP.

XX AAD31836;  
 AC AAD31836;

XX 18-JUN-2002 (first entry)  
 DT Human pancreatic tumour protein encoding cDNA clone 53803.2.

XX Human; pancreatic tumour protein; pancreatic cancer; therapy; diagnosis;  
 KW cancer; vaccine; cytostatic; ss.  
 KW Homo sapiens.

OS WO200212327-A2.

PN 14-FEB-2002.

XX 30-JUL-2001; 2001WO-US23969.

XX 03-AUG-2000; 2000US-223102P.

PR



QY 301 TTCCGAGAGGAGCCTCAGAAACGGCTACCAATCCCAAGGAGGACGCGCGCAA 360  
|||||  
DB 516 TTCCGAGAGGAGCCTCAGAAACGGCTACCAATCCCAAGGAGGACGCGCGCAA 457  
|||||  
QY 361 TTACCCACTCCG 373  
|||||  
DB 456 TTACCCACTCCG 444

## RESULT 11

AA06214/c  
ID AAL06214 standard; DNA; 13076 BP.  
XX  
AC AAL06214;  
XX  
XX 21-NOV-2001 (first entry)  
XX Human reproductive system related antigen DNA SEQ ID NO: 8902.  
DE Human reproductive system related antigen; reproductive system disorder;  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX Homo sapiens.  
OS  
XX WO200155320-A2.  
PN  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
XX 31-JAN-2000; 2000US-0175065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.











```
QY 61 GCTGACCCCTTCGCGGGGGGATGCGTGCATTATCAGATCAAGACCAACCGGTCAGC 120
Db 187 GCTGACCCCTTCGCGGGGGGATGCGTGCATTATCAGATCAAAACCAACCGGTCAGC 246
QY 121 CCTCTCGCGCCCGCGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 180
Db 247 CCTCTCGCGCCCGCGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 306
QY 181 GGGCGGATCGACACCCCGCTGGCGGCGAGACCCATTGCAAGCTCTGCCCTATCAACT 240
Db 307 GGGCGGATCGACACCCCGCTGGCGGCGAGACCCATTGCAAGCTCTGCCCTATCAACT 366
QY 241 TTCGATGTAAGTCGCCGCTGCTACCATGGTGACACCGGTTGACGGGAATCAGGGTTGCA 300
Db 367 TTCGATGTAAGTCGCCGCTGCTACCATGGTGACACCGGTTGACGGGAATCAGGGTTGCA 426
QY 301 TTCGGGAGAGGACCTTGAGAACGGCTACACATCCAAAGGAGGAGGCGCGGCAAA 360
Db 427 TTCGGGAGAGGACCTTGAGAACGGCTACACATCCAAAGGAGGAGGCGCGGCAAA 486
QY 361 TTACCCACTCCCG 373
Db 487 TTACCCACTCCCG 499

RESULT 15
ID AAD14299 standard; DNA; 1869 BP.
XX AAD14299;
XX
XX 06-NOV-2001 (first entry)
XX Human DNA to infer 18S ribosomal RNA.
DE
KW Transcriptional regulatory element; translational regulatory element;
KW gene expression; diagnostic application; therapeutic application;
KW human; ribosomal RNA; 18S rRNA; internal ribosome entry site; IRES; ds.
XX
XX Homo sapiens.
XX
FH key Location/Qualifiers
FT modified_base 27
FT /*tag= a
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 99
FT /*tag= b
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 116
FT /*tag= c
FT /*mod_base= um
FT modified_base 121
FT /*tag= d
FT /*mod_base= um
FT modified_base 159
FT /*tag= e
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 166
FT /*tag= f
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 172
FT /*tag= g
FT /*mod_base= um
FT modified_base 174
FT /*tag= h
FT /*mod_base= cm
FT modified_base 428
FT /*tag= i
FT /*mod_base= um

modified_base 436
FT /*tag= j
FT /*mod_base= gm
modified_base 462
FT /*tag= k
FT /*mod_base= cm
modified_base 468
FT /*tag= l
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 484
FT /*tag= m
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 509
FT /*tag= n
FT /*mod_base= gm
modified_base 512
FT /*tag= o
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 517
FT /*tag= p
FT /*mod_base= cm
modified_base 576
FT /*tag= q
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 590
FT /*tag= r
FT /*mod_base= gm
modified_base 601
FT /*tag= s
FT /*mod_base= gm
modified_base 627
FT /*tag= t
FT /*mod_base= um
modified_base 644
FT /*tag= u
FT /*mod_base= gm
modified_base 668
FT /*tag= v
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 683
FT /*tag= w
FT /*mod_base= gm
modified_base 797
FT /*tag= x
FT /*mod_base= cm
modified_base 799
FT /*tag= y
FT /*mod_base= um
modified_base 867
FT /*tag= z
FT /*mod_base= gm
modified_base 1031
FT /*tag= aa
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
modified_base 1248
FT /*tag= ab
FT /*mod_base= OTHER
FT /*note= "3-(3-amino-3-carboxypropyl)-
1-methylpseudouridine"
modified_base 1288
FT /*tag= ac
FT /*mod_base= um
modified_base 1326
FT /*tag= ad
FT /*mod_base= um
modified_base 1328
FT
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:47:14 ; Search time 52 Seconds  
(without alignments)  
963.781 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQGGQQPQGGQGG.....KSLVQIHEKNGWYTPPKEDG 422

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1863	83.6	374	15	US-10-106-698-4458
2	140.5	6.3	46	9	US-09-864-761-43956
3	137	6.1	165	9	US-09-842-528-4
4	135.5	6.1	295	15	US-10-060-019-22
5	134.5	6.0	223	15	US-10-128-714-8332
6	133	6.0	238	15	US-10-128-714-3352
7	122.5	5.5	1048	15	US-10-174-363-56
8	121.5	5.5	144	9	US-09-842-528-5
9	121.5	5.5	1046	15	US-10-156-761-10088
10	120.5	5.4	1004	10	US-09-738-626-5676
11	119.5	5.4	256	9	US-09-925-301-1083
12	118	5.3	527	10	US-09-738-973-216
13	118	5.3	527	10	US-09-854-133-216
14	118	5.3	527	15	US-10-144-649A-216
15	116.5	5.2	1786	10	US-09-742-096-3

16	116	5.2	1647	11	US-09-824-574-4	Sequence 4, Appli
17	115.5	5.2	265	11	US-09-975-719-31	Sequence 31, Appli
18	114	5.1	519	15	US-10-113-794A-2	Sequence 2, Appli
19	113.5	5.1	176	15	US-10-043-487-401	Sequence 401, App
20	112.5	5.0	161	9	US-09-925-297-801	Sequence 801, App
21	111.5	5.0	147	15	US-10-153-668-292	Sequence 292, App
22	111.5	5.0	147	15	US-10-153-668-406	Sequence 406, App
23	111.5	5.0	525	15	US-10-168-425-7	Sequence 7, Appli
24	109.5	4.9	147	14	US-10-108-605-331	Sequence 331, App
25	109	4.9	1084	15	US-10-217-700-9	Sequence 9, Appli
26	108	4.8	592	14	US-10-047-593-2	Sequence 2, Appli
27	108	4.8	592	14	US-10-047-593-4	Sequence 4, Appli
28	107.5	4.8	147	15	US-10-153-668-294	Sequence 294, App
29	107.5	4.8	147	15	US-10-153-668-408	Sequence 408, App
30	107.5	4.8	148	9	US-09-826-312-4	Sequence 4, Appli
31	107.5	4.8	148	15	US-10-108-767-4	Sequence 4, Appli
32	107.5	4.8	148	15	US-10-152-156-4	Sequence 4, Appli
33	107	4.8	156	15	US-10-043-487-412	Sequence 412, App
34	106.5	4.8	143	15	US-10-043-487-399	Sequence 399, App
35	106.5	4.8	641	11	US-09-959-987-2	Sequence 2, Appli
36	106.5	4.8	647	9	US-09-925-299-1002	Sequence 1002, Ap
37	106.5	4.8	647	11	US-09-925-299-1002	Sequence 392, App
38	105.5	4.7	143	15	US-10-043-487-392	Sequence 324, App
39	105.5	4.7	462	11	US-09-919-039-324	Sequence 8314, Ap
40	105.5	4.7	1070	15	US-10-128-714-8314	Sequence 56, Appli
41	105	4.7	510	9	US-09-866-562-56	Sequence 150, App
42	104.5	4.7	151	9	US-09-842-528-2	Sequence 2, Appli
43	104.5	4.7	151	15	US-10-153-668-150	Sequence 2, Appli
44	103.5	4.6	758	10	US-09-903-248-2	Sequence 2, Appli
45	103.5	4.6	758	10	US-09-859-604-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-106-698-4458

; Sequence 4458, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept

; FILE REFERENCE: P0005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; PRIOR FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn Ver. 3.0

; SEQ ID NO 4458

; LENGTH: 374

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: MISC\_FEATURE

; LOCATION: (25)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: MISC\_FEATURE

; LOCATION: (38)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-106-698-4458

Query Match 83.6%; Score 1863; DB 15; Length 374;

Best Local Similarity 98.4%; Pred. No. 3.1e-132;

Matches 359; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 59 FRTASACLDSECEFLLAGAGAGAGAPGPHLPGRGSPGDPVRTHCNITESYPAPVPI 118

Db 11 FRTASACLDSECEFLLAGAGAGAGAXPGTASPTGSPGDPVRTHCNITESYPAPVPI 70









QY 96 -----SVPGDPVR-----IHCNITESPAY--PPIWSVEDDPNLAAY-LE- 134  
Db 400 PGGGGTQAGFGRPGARRGKSKRRQREVEAQAQSVGVMPLPRNGQSVRLSRGAS 459  
QY 135 LVDIKKGTLLQHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDYSSDEDEEM 194  
Db 460 LTDFAEKIGANPASLVGMNMLGEMVTATQSVSDTLKLLADEMNFVLEIVSPEEDREL 519  
QY 195 PEDTDLHYEMKEEPAE-----CKKSEDDGIGKENLAILEIKKN 236  
Db 520 LESFD-----IFEDEGEDEELVAPPPVTVMGHVDHGKTRLLDTIRKTNVAGEA----- 571  
QY 237 QRODYNGAVS---GSVQATDRMLKELRDI-----YRSQFKGNGVAVELVN 280  
Db 572 -----GGIQTGHGAYQVTFVNDDEERKIFIDTPGHEATAMRARGAKSTDAI----- 620  
QY 281 DSLIDYNNVKKLVQDSDALHNDLQILKEKGADE-ILLNFSGKDNFPDPFVRYVSPVL 339  
Db 621 -----LVVAANDGYMPOTIEALNHAKAADVPVAVNKNIDVEGADPTKVR--GOLT 669  
QY 340 SGGYV---LGGGAICMELLTKQSSAYSIESVIMQISATL 377  
Db 670 EFGVAAEYGGTMEFVDSIAKQGLNIESILLEAVLTDASL 710

RESULT 10

US-09-738-626-5676  
; Sequence 5676, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 5676  
; LENGTH: 1004  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5676

Query Match 5.4%; Score 120.5; DB 10; Length 1004;  
Best Local Similarity 21.7%; Pred. No. 1.6;  
Matches 97; Conservative 46; Mismatches 112; Indels 191; Gaps 24;  
QY 4 PQPGQQQPGGQQLGGGGA-----AP---GAGGGPGG 33  
Db 281 PRPGGSRPGGSAQAQAQPSQERQGGRRPSPAMPPPTQCPAKAPGKGRGQAGG 340  
QY 34 GPGGPGCLRLKLLLESIFHRGHERFIASACLDLSCEFFLAGGAGAGAAACPHLPP 93  
Db 341 GAGGG-----FNRG-----GGTGGGAGRGR--- 361  
QY 94 RGSVPG-----DPVRHCNITESPAYPPPIWSVEDDPNLAAYLERLVDIKK 140  
Db 362 RGGTAGAGRPGGAPRRGKSKRRQREYESNQA-----PNVIGGV-RLPD-GK 408

QY 141 GNTL-----LLQHLKRIISDLCKLYNLPQ--HPDVEMLDQPLPAEQCTQED--- 184  
Db 409 GATIRLAGASLADFADKIGADAAALVOALFNLGEMVTATASVSDETL---QLLGEEMNY 465  
QY 185 ---VSSEDEDEEMPEDETLDDHYEMKEEPAEKGKSEDDGIGKENLA----- 228  
Db 466 KVQVVSPEDEREL-----LESFDL-----QFGEDEG-GEADLAKRPPVTVMGHV 510  
QY 229 -----ILEKIKK-NORQDYNGAVSG-----SVOATDRML-----KELRDIYRS 266  
Db 511 DHGKTRLLDTIRKANVGSDEAGGITQIGAYQVKNVVEDTERTITFLDTGHEAFTAMRA 570  
QY 267 QSFKGGNYAVELV--NDSLYDWNVKKL-----KVDQDSALHNDLQ----- 304  
Db 571 RGAKTSTDAVLVWAADDGVMPTVEALNHAKAADVPVAVNKNIDKPEASPEKIRGOLTE 630  
QY 305 --ILKEKEGADFILLNFSFKDNFPFD 328  
Db 631 YGLIPEEYGGDTIFVDSIAKQGLNID 656

RESULT 11

US-09-925-301-1083  
; Sequence 1083, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: Patent in ver. 2.0  
; SEQ ID NO 1083  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (56)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-1083

Query Match 5.4%; Score 119.5; DB 9; Length 256;  
Best Local Similarity 22.9%; Pred. No. 0.29;  
Matches 49; Conservative 32; Mismatches 96; Indels 37; Gaps 8;  
QY 187 SEDEDEEMPEDETLDDHYEMKEEPAEKGKSEDDGIGKENLAILEIKKNQDQYNGAV 246  
Db 65 SDDDSRASTSSSSSSNQOTEKETNTPKKKESKSKSKLLSTSAK----- 113  
QY 247 SGSVQATDRMLKELRDIYRSQSFKGGNYAVELVNDLSLYDNVKKLVQDQDSALHNDLQIL 306  
Db 114 -----RIQELADITLDP-----PNCAGPKGDNIEWRSTIL--GPGSVY----- 154  
QY 307 KEKGADEPILLNFSKDNFPDPFVRYVSPVLSGGVVLGGGGAICMELLTKQGNSSAYS 366  
Db 155 ---EGGYF-FIDITFTPEYFPKPKVTRFTRIYHCN-INSQGVICLDIL-KDNWSPALTI 208  
QY 367 ESVMIOISATLVK-GKARVQFGANKSOYSLTRAQ 399  
Db 209 SKVLISCSLLTDCNPADPLVGSITATQYMTNRAE 242

RESULT 12

US-09-738-973-216  
; Sequence 216, Application US/09738973  
; Patent No. US20020110563A1

[illegible]





RESULT 3  
US-08-093-144-6  
; Sequence 6, Application US/08093144  
; Patent No. 5434048  
; GENERAL INFORMATION:  
; APPLICANT: SIMON, LUC  
; APPLICANT: LAIONDE, MAURICE  
; TITLE OF INVENTION: DNA PROBES FOR THE DETECTION OF  
; TITLE OF INVENTION: ARBUSCULAR ENDOMYCORRHIZAL FUNGI-  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

```

RESULT 4
US-08-998-416-952/c
: Sequence 952, Application US/08998416
: Patent No. 6239264
: GENERAL INFORMATION:
: APPLICANT: Philippson, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Nechtelle, Philipp
: APPLICANT: Rebschung, Corinne
: TITLE OF INVENTION: GENOMIC DNA S
: TITLE OF INVENTION: AND USES THERE
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 6239264artis Cor
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park

```

```

; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 952:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1588UP
;
; US-08-998-416-952
;
; Query Match
; Best Local Similarity 6.1%; Score 137.2; DB 3; Length 723;
; Matches 169; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
;
; QY 150 GCCCGCGGCTTTGGTACTCTAGATACCTCGGGCGGATCGACGCCGCCCGTGGCGGG 209
; Db 454 GTCTTCGGACTCTTTGATGATTCAATAAATTTTCGAATCGATGCTGTGCTGGCG 395
;
; QY 210 AGCACCATTGCAACGCTCTGCTTCACTTTCGATGCTAGTCCCGTGCCTACCATGG 269
; Db 394 ATGGTTCATCAATTTCTGCCCTATCAACTTTCGATGCTAGTCCCGTGCCTACCATGG 335
;
; QY 270 TGACCACGGGTGACGGGGAATCAGGTTTCGATTCGGAGAGGAGGAGGAGGAGGAGG 329
; Db 334 TTTCACGGGTAAACGGGAATAAGGTTTCGATTCGGAGAGGAGGAGGAGGAGGAGG 275
;
; QY 330 CCACATCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 371
; Db 274 CCACATCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 233
;
; RESULT 5
; US-09-350-710B-1/c
; Sequence 1, Application US/09350710B
; Patent No. 6489158
; GENERAL INFORMATION:
; APPLICANT: Hendrick, Carol A.
; APPLICANT: Platt, Nancy J.
; APPLICANT: Ruser, Barbara G.
; APPLICANT: Hoganson, Dean A.
; TITLE OF INVENTION: No. 6489158-Lactate Assimilating Yeast for Improving Aerobic
; FILE REFERENCE: Stability of Silage
; CURRENT APPLICATION NUMBER: US/09/350,710B
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1776
; TYPE: DNA
;
; ORGANISM: Yeast integration vector
; US-09-350-710B-1
;
; Query Match
; Best Local Similarity 6.1%; Score 137.2; DB 4; Length 1776;
; Matches 169; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
;
; QY 150 GCCCGCGGCTTTGGTACTCTAGATACCTCGGGCGGATCGACGCCGCCCGTGGCGGG 209
; Db 1550 GTCTTCGGACTCTTTGATGATTCAATAAATTTTCGAATCGATGCTGTGCTGGCG 1491
;
; QY 210 AGCACCATTGCAACGCTCTGCTTCACTTTCGATGCTAGTCCCGTGCCTACCATGG 269
; Db 1490 ATGGTTCATTCAAATTTCTGCCCTATCAACTTTCGATGCTAGTCCCGTGCCTACCATGG 1431
;
; QY 270 TGACCACGGGTGACGGGGAATCAGGTTTCGATTCGGAGAGGAGGAGGAGGAGGAGG 329
; Db 1430 TTTCACGGGTAAACGGGAATAAGGTTTCGATTCGGAGAGGAGGAGGAGGAGGAGG 1371
;
; QY 330 CCACATCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 371
; Db 1370 CCACATCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1329
;
; RESULT 6
; US-08-998-416-178/c
; Sequence 178, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippson, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgin
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 178:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

```



; ORGANISM: PAG1070RP  
US-08-998-416-178

Query Match 6.1%; Score 136.8; DB 3; Length 703;  
Best Local Similarity 75.7%; Pred. No. 4.7e-17;  
Matches 168; Conservative 1; Mismatches 53; Indels 0

[illegible]

## RESULT. T 7

US-08-998-416-281/C  
Sequence 281, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgен  
APPLICANT: Knechtle, Philipp  
APPLICANT: Rebschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

US-08-998-416-281  
; ORGANISM: PAG1237UP

```
Query Match      6.1%; Score 136.6; DB 3; Length 709;
Best Local Similarity 86.3%; Pred. No. 5.1e-17;
Matches 151; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

197	CCCCGTGGCGGACGACGCCCATTCGAAACGCTCTGCCCTATCAACTTTTCGATGGTAGTCGCC	256
QY		
407	CTTGTGCTGGGATGGTTTCATTCAAATTTCTGCCCTATCAACTTTTCGATGGTAGGATAG	348
Db		
257	GTGCCTACCATGGTGTGACACGGGTGACGGGGAATCAGGGTTCGATTCGCGAGAGGGAGCC	316
QY		
347	TGGCCTACCATGGTTTCAACGGGTACGGGGAATAAGGGTTCGATTCGCGAGAGGGAGCC	388
Db		
317	TGAGAAACGGGTACACATCCACAGGAAGGACGAGCGCGGCAAAATTACCCACTCC	371
QY		
287	TGAGAAACGGGTACCATTCACAGGAAGGACGAGCGCGGCAAAATTACCCAACTCC	233
Db		

## RESULT 8

```

US-08-967-101-19/C
Sequence 19, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-967-101-19

```

	Query Match	6.08;	Score 135.4;	DB 2;	Length 395;
	Best Local Similarity	98.7%;	Pred. No. 7.3e-17;		
	Matches 147;	Conservative	0;	Mismatches 1;	Indels 1; Gaps 1
QY	226	TCGCGCCTATCAACTTTCGATGTTAGTCGCGCGTGCCT-ACCATGTTGACCAACGGGTGACG	284		
Db	180	TCGCGCCTATTAACTTTTCGATGTTAGTCGCGCGTGCCTAAACATGTTGACCAACGGGTGACG	121		
OY	285	GGGAATCAGGGTTCGATTTCGCGAGAGGGAGCGCTCAGAAACGGCTACACATCTCCAAAGGAG	344		

Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAACGGCTACCATCCACGGAAG 61  
Qy 345 GCAGCAGCGCGCAAAATACCACCTCCG 373  
Db 60 GCAGCAGCGCGCAAAATACCACCTCCG 32

## RESULT 9

US-08-592-541-19/c  
; Sequence 19, Application US/08592541  
; Patent No. 5986054  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/592,541  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-592-541-19

Query Match 6.0%; Score 135.4; DB 2; Length 395;  
Best Local Similarity 98.7%; Pred. No. 7.3e-17;  
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 226 TCTGCCCTATCAACTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 284  
Db 180 TCTGCCCTATTAACCTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 121  
Qy 285 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 344  
Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 61  
Qy 345 GCAGCAGCGCGCAAAATACCACCTCCG 373  
Db 60 GCAGCAGCGCGCAAAATACCACCTCCG 32

## RESULT 10

US-09-124-698-19/c  
; Sequence 19, Application US/09124698  
; Patent No. 6117978  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M

; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/09/124,698  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/592,541  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-124-698-19  
Query Match 6.0%; Score 135.4; DB 3; Length 395;  
Best Local Similarity 98.7%; Pred. No. 7.3e-17;  
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 226 TCTGCCCTATCAACTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 284  
Db 180 TCTGCCCTATTAACCTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 121  
Qy 285 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 344  
Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCACGGAAG 61  
Qy 345 GCAGCAGCGCGCAAAATACCACCTCCG 373  
Db 60 GCAGCAGCGCGCAAAATACCACCTCCG 32

## RESULT 11

US-09-127-480-19/c  
; Sequence 19, Application US/09127480  
; Patent No. 6194153  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:



Query Match 6.0%; Score 135.4; DB 4; Length 395;  
Best Local Similarity 98.7%; Pred. No. 7.3e-17;  
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTTCGATGTAGTCGCCGTGCCT-ACCATGTGACCAAGGGGTGACG 284  
|||||  
Db 180 TCTGCCCTATTAACTTTCGATGTAGTCGCCGTGCCTAACCATGTGACCAAGGGGTGACG 121  
|||||  
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 344  
|||||  
Db 120 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 61  
|||||  
QY 345 GCAGCAGGCGCGCAAAATTACCCACTCCCG 373  
|||||  
Db 60 GCAGCAGGCGCGCAAAATTACCCACTCCCG 32  
|||||

## RESULT 14

US-09-636-796A-19/c  
; Sequence 19, Application US/09636796A  
; Patent No. 6485911

## GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; ROMMENS, JOHANNA M  
; FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/636,796A  
FILING DATE: 11-Aug-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-636-796A-19

Query Match 6.0%; Score 135.4; DB 4; Length 395;  
Best Local Similarity 98.7%; Pred. No. 7.3e-17;  
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTTCGATGTAGTCGCCGTGCCT-ACCATGTGACCAAGGGGTGACG 284  
|||||  
Db 180 TCTGCCCTATTAACTTTCGATGTAGTCGCCGTGCCTAACCATGTGACCAAGGGGTGACG 121  
|||||  
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 344  
|||||  
Db 120 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 61  
|||||

QY 345 GCAGCAGGCGCGCAAAATTACCCACTCCCG 373  
|||||  
Db 60 GCAGCAGGCGCGCAAAATTACCCACTCCCG 32  
|||||

## RESULT 15

US-08-431-048F-19/c

; Sequence 19, Application US/08431048F

; Patent No. 6531586

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; ROMMENS, JOHANNA M  
; FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 155

CORRESPONDENCE ADDRESS:

ADDRESSEE: DARBY & DARBY P.C.

STREET: 805 THIRD AVENUE

CITY: NEW YORK

STATE: N.Y.

COUNTRY: U.S.A.

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/431,048F

FILING DATE: 28-Apr-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: FEHLNER, PAUL F.

REGISTRATION NUMBER: 35135

REFERENCE/DOCKET NUMBER: 1034/0F808

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-527-6237

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-08-431-048F-19

Query Match 6.0%; Score 135.4; DB 4; Length 395;  
Best Local Similarity 98.7%; Pred. No. 7.3e-17;  
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTTCGATGTAGTCGCCGTGCCT-ACCATGTGACCAAGGGGTGACG 284  
|||||  
Db 180 TCTGCCCTATTAACTTTCGATGTAGTCGCCGTGCCTAACCATGTGACCAAGGGGTGACG 121  
|||||  
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 344  
|||||  
Db 120 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCCTGAGAAACGGCTACCAATCCAAAGGAG 61  
|||||  
QY 345 GCAGCAGGCGCGCAAAATTACCCACTCCCG 373  
|||||  
Db 60 GCAGCAGGCGCGCAAAATTACCCACTCCCG 32  
|||||

Search completed: August 4, 2003, 19:07:44  
Job time : 199 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 16:06:37 ; Search time 652 Seconds  
(without alignments)  
7131.908 Million cell updates/sec

Title: US-10-005-549-1  
Perfect score: 2254  
Sequence: 1 ctcgctctctctactgg.....tttgaaaaa.....2254

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1554.4	69.0	1666	14	US-10-106-698-181
C 2	432	19.2	432	10	US-09-880-107-3600
C 3	371.4	16.5	870	10	US-09-919-344-14
C 4	371.4	16.5	13076	11	US-09-764-891-8902
C 5	371.4	16.5	13076	11	US-09-764-891-10043
C 6	371.4	16.5	42999	10	US-09-799-462A-17
C 7	371.4	16.5	42999	11	US-09-836-911A-17
C 8	371.4	16.5	42999	13	US-10-125-767-17
C 9	371.4	16.5	42999	14	US-10-151-081-17
C 10	371.4	16.5	42999	14	US-10-287-313-17
C 11	371.4	16.5	42999	14	US-10-219-694-17
C 12	368.2	16.3	1969	10	US-09-880-107-2231
C 13	360	16.0	427	11	US-09-918-995-19104
C 14	352.2	15.6	416	11	US-09-918-995-5393
C 15	349.8	15.5	393	11	US-09-918-995-36028
C 16	333.4	14.8	343	10	US-09-796-692-8493

C 17	333.4	14.8	343	14	US-10-040-862-8493	Sequence 8493, Ap
C 18	333	14.8	536	14	US-10-106-698-2181	Sequence 2181, Ap
C 19	331.4	14.7	17384	14	US-10-161-403-118	Sequence 118, Ap
C 20	331.4	14.7	22118	10	US-09-799-462A-16	Sequence 16, Ap
C 21	331.4	14.7	22118	11	US-09-815-981-5	Sequence 5, Ap
C 22	331.4	14.7	22118	11	US-09-836-911A-16	Sequence 16, Ap
C 23	331.4	14.7	22118	11	US-09-815-979-5	Sequence 5, Ap
C 24	331.4	14.7	22118	13	US-10-125-767-16	Sequence 16, Ap
C 25	331.4	14.7	22118	14	US-10-151-081-16	Sequence 16, Ap
C 26	331.4	14.7	22118	14	US-10-287-313-16	Sequence 16, Ap
C 27	331.4	14.7	22118	14	US-10-219-694-16	Sequence 16, Ap
C 28	331.4	14.7	22118	14	US-10-235-119-5	Sequence 5, Ap
C 29	331.4	14.7	22118	14	US-10-161-403-18	Sequence 18, Ap
C 30	312.6	13.9	696	14	US-10-198-846-9414	Sequence 9414, Ap
C 31	290.8	12.9	488	11	US-09-918-995-17863	Sequence 17863, A
C 32	281	12.5	281	10	US-09-998-598-2390	Sequence 2390, Ap
C 33	274.4	12.2	374	13	US-10-079-623-137	Sequence 137, Ap
C 34	269.4	12.0	316	10	US-09-880-107-194	Sequence 194, Ap
C 35	262	11.6	510	11	US-09-918-995-16742	Sequence 16742, A
C 36	257.4	11.4	792	13	US-10-027-632-28550	Sequence 28550, A
C 37	227.4	10.1	357	10	US-09-867-701-2408	Sequence 2408, Ap
C 38	222.8	9.9	822	14	US-10-198-846-9042	Sequence 9042, Ap
C 39	220	9.8	220	10	US-09-878-178-541	Sequence 541, Ap
C 40	220	9.8	220	13	US-10-046-935-541	Sequence 541, Ap
C 41	220	9.8	220	14	US-10-146-502-541	Sequence 541, Ap
C 42	220	9.8	378	10	US-09-799-462A-21	Sequence 21, Ap
C 43	220	9.8	378	11	US-09-815-981-9	Sequence 9, Ap
C 44	220	9.8	378	11	US-09-836-911A-21	Sequence 21, Ap
C 45	220	9.8	378	11	US-09-815-979-9	Sequence 9, Ap

ALIGNMENTS

RESULT 1

US-10-106-698-181/c  
; Sequence 181, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10106.698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 181  
; LENGTH: 1666  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (8)..(8)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc.feature  
; LOCATION: (1553)..(1553)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-181

Query Match 69.0%; Score 1554.4; DB 14; Length 1666;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1572; Conservative 1; Mismatches 13; Indels 1; Gaps 1;  
Qy 661 GAGTCCATCTTCCACCGGCCAGGCG-CTTCCGATTCACAGCCCTTCCTGGACGA 719  
Db 1666 GAGTCCATCTTCCACCGGCCAGGCGCTTCCGATTCACAGCCCTTCCTGGACGA 1607





```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5583)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5606)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (11536)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-891-8902
```

```
Query Match 16.5%; Score 371.4; DB 11; Length 13076;
Best Local Similarity 99.7%; Pred. No. 4e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGCTCCTCTCCTTGGTAACTGTGGTAATTCYAGAGCTAATACATGCCGACGGC 60
Db 3632 CTCGCTCCTCTCCTTGGTAACTGTGGTAATTCYAGAGCTAATACATGCCGACGGC 3573

Qy 61 GCTGACCCCTTCGCGGGGGGATGCGTGATTTATCAGATCAAGACCAACCCGGTCAGC 120
Db 3572 GCTGACCCCTTCGCGGGGGGATGCGTGATTTATCAGATCAAGACCAACCCGGTCAGC 3513

Qy 121 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGGCTTGGTGACTCTAGATAACCTC 180
Db 3512 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGGCTTGGTGACTCTAGATAACCTC 3453

Qy 181 GGGCGGATGCGACGCCCGCGCGCGCGGAGCCATTCGAAGCTCTGCCCTATCAACT 240
Db 3452 GGGCGGATGCGACGCCCGCGCGCGGAGCCATTCGAAGCTCTGCCCTATCAACT 3393

Qy 241 TTCGATGGTAGTCCCGTCCCTACCATGGTGACACCGGGTGACGGGGAATCAGGGTTCGA 300
Db 3392 TTCGATGGTAGTCCCGTCCCTACCATGGTGACACCGGGTGACGGGGAATCAGGGTTCGA 3333

Qy 301 TTCGGGAGAGGAGCCCTGAGAAACGGCTACACATCCAAAGGAGGCGGCGCAAA 360
Db 3332 TTCGGGAGAGGAGCCCTGAGAAACGGCTACACATCCAAAGGAGGCGGCGCAAA 3273

Qy 361 TTACCCACTCCG 373
Db 3272 TTACCCACTCCG 3260
```

## RESULT 5

```
US-09-764-891-10043/c
; Sequence 10043, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10043
; LENGTH: 13076
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1259)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1260)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

```
; LOCATION: (1261)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1262)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1263)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1311)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1313)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (4089)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5583)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5606)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (11536)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-891-10043
```

```
Query Match 16.5%; Score 371.4; DB 11; Length 13076;
Best Local Similarity 99.7%; Pred. No. 4e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGCTCCTCTCCTTGGTAACTGTGGTAATTCYAGAGCTAATACATGCCGACGGC 60
Db 3632 CTCGCTCCTCTCCTTGGTAACTGTGGTAATTCYAGAGCTAATACATGCCGACGGC 3573

Qy 61 GCTGACCCCTTCGCGGGGGGATGCGTGATTTATCAGATCAAGACCAACCCGGTCAGC 120
Db 3572 GCTGACCCCTTCGCGGGGGGATGCGTGATTTATCAGATCAAGACCAACCCGGTCAGC 3513

Qy 121 CCCTCTCCGGCGCGCGGGGGGCGCGCGCGGCTTGGTGACTCTAGATAACCTC 180
Db 3512 CCCTCTCCGGCGCGCGGGGGGCGCGCGCGGCTTGGTGACTCTAGATAACCTC 3453

Qy 181 GGCGCGATCGACGCCCGCGCGCGCGGCTTGGTGAGACCCATTCGAAGCTCTGCCCTATCAACT 240
Db 3272 TTACCCACTCCG 3260
```



Db 3452 GGGCGGATCGCAGCGCCCGCTGGCGGCGAGCACCATTTCGAACGCTGCGCCCTATCAACT 3393  
QY 241 TTCGATGTAGTCGCGGCTCCCTACCATGTGTGACCGGCTGACGGGGAATCAGGGTTGCA 300  
Db 3392 TTCGATGTAGTCGCGGCTCCCTACCATGTGTGACCGGCTGACGGGGAATCAGGGTTGCA 3333  
QY 301 TTCGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAATCCCAAGGAGCAGCGCGCAAA 360  
Db 3332 TTCGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAATCCCAAGGAGCAGCGCGCAAA 3273  
QY 361 TTACCCACTCCCG 373  
Db 3272 TTACCCACTCCCG 3260

## RESULT 6

US-09-799-462A-17  
; Sequence 17, Application US/09799462A  
; Patent No. US20020160970A1

## GENERAL INFORMATION:

APPLICANT: Hadlaczky, Gyula  
Szalay, Aladar

## TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF

AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

## NUMBER OF SEQUENCES: 34

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe

STREET: 4250 Executive Square, 7th Floor

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/799,462A

FILING DATE: 10-Sep-2001

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,682

FILING DATE: 10-APR-1997

APPLICATION NUMBER: 08/695,191

FILING DATE: 07-AUG-1996

APPLICATION NUMBER: 08/682,080

FILING DATE: 15-JUL-1996

APPLICATION NUMBER: 08/629,822

FILING DATE: 10-APR-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24601-402G

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 858-450-8403

TELEFAX: 858-587-5360

TELEX: <Unknown>

## INFORMATION FOR SEQ ID NO: 17:

## SEQUENCE CHARACTERISTICS:

LENGTH: 42998 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

## FRAGMENT TYPE: &lt;Unknown&gt;

ORIGINAL SOURCE:

SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-09-799-462A-17

Query Match

16.5%; Score 371.4; DB 10; Length 42999;

Best Local Similarity 99.7%; Pred. No. 8.5e-87;  
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCTCCTCTCCTACTTGGATAACTGTGTGTAATTTCTAGAGCTAATACATGCCGACGGCG 60  
Db 3783 CTCGCTCCTCTCCTACTTGGATAACTGTGTGTAATTTCTAGAGCTAATACATGCCGACGGCG 3842  
QY 61 GCTGACCCCTTTCGGGGGGGGGATCGGTGCAATTTATCAGATCAAGACCAACCCCGGTGACG 120  
Db 3843 GCTGACCCCTTTCGGGGGGGGGATCGGTGCAATTTATCAGATCAAAACCAACCCCGGTGACG 3902  
QY 121 CCCTCTCGGCGCCCGGG 180  
Db 3903 CCCTCTCGGCGCCCGGG 3962  
QY 181 GGGCGGATCGCAGCGCCCGCTGGCGGCGAGCAGCCATTTCGAAGCTCTGCCCTATCAACT 240  
Db 3963 GGGCGGATCGCAGCGCCCGCTGGCGGCGAGCAGCCATTTCGAAGCTCTGCCCTATCAACT 4022  
QY 241 TTCGATGTAGTCGCGGCTTACCATGCTGACCATGCTGACCATGCTGACCATGCTGACCATG 300  
Db 4023 TTCGATGTAGTCGCGGCTTACCATGCTGACCATGCTGACCATGCTGACCATGCTGACCATG 4082  
QY 301 TTCGAGAGGAGGAGCTGAGAAACGGCTACCATGCTGACCATGCTGACCATGCTGACCATG 360  
Db 4083 TTCGAGAGGAGGAGCTGAGAAACGGCTACCATGCTGACCATGCTGACCATGCTGACCATG 4142  
QY 361 TTACCCACTCCCG 373  
Db 4143 TTACCCACTCCCG 4155

## RESULT 7

US-09-836-911A-17  
; Sequence 17, Application US/09836911A  
; Publication No. US20030033617A1

## GENERAL INFORMATION:

APPLICANT: Hadlaczky, Gyula

Szalay, Aladar

## TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF

AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

## NUMBER OF SEQUENCES: 34

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe

STREET: 4350 La Jolla Village Drive, 6th Floor

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92122

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/836,911A

FILING DATE: 17-Apr-2002

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,682

FILING DATE: 10-APR-1997

APPLICATION NUMBER: 08/695,191

FILING DATE: 07-AUG-1996

APPLICATION NUMBER: 08/682,080

FILING DATE: 15-JUL-1996

APPLICATION NUMBER: 08/629,822

FILING DATE: 10-APR-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24601-402I

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 858-450-8403

/		TELEFAX: 858-587-5360	
/		TELEX: <Unknown>	
/		INFORMATION FOR SEQ ID NO: 17:	
/		SEQUENCE CHARACTERISTICS:	
/		LENGTH: 42999 base pairs	
/		TYPE: nucleic acid	
/		STRANDEDNESS: single	
/		TOPOLOGY: linear	
/		MOLECULE TYPE: Genomic DNA	
/		HYPOTHETICAL: NO	
/		ANTI-SENSE: NO	
/		FRAGMENT TYPE: <Unknown>	
/		ORIGINAL SOURCE:	
/		SEQUENCE DESCRIPTION: SEQ ID NO: 17:	
/		US-09-836-911A-17	
Query Match                      16.5%; Score 371.4; DB 11; Length 42999;			
Best Local Similarity 99.7%; Pred. No. 8.5e-87;			
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	CTCGCCTCTCTACTTGGATAACGTGCTTAATTTAGAGCTAAATACATGCCGAGGGGC	60
Dd	3783		3842
QY	61	GCTGACCCCTTCGCGGGGGGATGCGTGATTATCAGATCAAGACCAACC CGGT CAGC	120
Dd	3843		3902
QY	121	CCCTCTCGGCCCCCGCGGGGGCGGCGCCCGCGCTTTGTGTGACTCTAGATAACCTC	180
Dd	3903	CCCTCTCGGCCCCCGCGGGGGCGGCGCCCGCGCTTTGTGTGACTCTAGATAACCTC	3962
QY	181	GGSCCGATGCACGCCGCCCGGTGCGSGCGACGACCATTCGGAACGCTCGCCCTACAAC	240
Dd	3963	GGSCCGATGCACGCCGCCCGGTGCGSGCGACGACCATTCGGAACGCTCGCCCTACAAC	4022
QY	241	TTCGATGGTAGTCGCGTGCGCTACCATGTGTGACACGGGTGACGGGSAATFCAGGGTTCGA	300
Dd	4023	TTCGATGGTAGTCGCGTGCGCTACCATGTGTGACACGGGTGACGGGSAATFCAGGGTTCGA	4082
QY	301	TTCGGGAGAGGGAGCCTGAGAAACGGCTACCATCCAGGAGGCGAGCGCGCAA	360
Dd	4083	TTCGGGAGAGGGAGCCTGAGAAACGGCTACCATCCAGGAGGCGAGCGCGCAA	4142
QY	361	TTACCACTCCCG 373	
Dd	4143	TTACCACTCCCG 4155	

RESULT 8.  
US-10-125-767-17  
; Sequence 17, Application US/10125767  
; Publication No. US20020160410A1  
; GENERAL INFORMATION:  
; APPLICANT: Hadlaczky, Gyula  
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
; METHODS  
; FOR PREPARING ARTIFICIAL CHROMOSOMES  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Heller Ehrman White & McAuliffe LLP  
; STREET: 4350 La Jolla Village Drive, 7th Floor  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/125.767

```

; FILING DATE: 17-Apr-2002
; PRIORITY: 1
; APPLICATION NUMBER: 09/724,693
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17

Query Match 16.5%; Score 371.4; DB 13; Length 42999;
Best Local Similarity 99.7%; Pred. No. 8.5e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCTCCTCTCTACTTGGATAAAGTGTGGTAATCTAGAGCTAATACATGCCGACGGGC 60
DB 3783 CTCGCTCCTCTCTACTTGGATAAAGTGTGGTAATCTAGAGCTAATACATGCCGACGGGC 3842

QY 61 GCTGACCCCTTCGCGGGGGGATGCGTGCATTTATCAGATCAAGACCAACCGGTTCAGC 120
DB 3843 GCTGACCCCTTCGCGGGGGGATGCGTGCATTTATCAGATCAAGACCAACCGGTTCAGC 3902

QY 121 CCTCTCCGCGCCCGCGCGGGGGCGCGCGCGCGCTTGTGTGACTCTAGATAACCTC 180
DB 3903 CCTCTCCGCGCCCGCGCGGGGGCGCGCGCGCGCTTGTGTGACTCTAGATAACCTC 3962

QY 181 GGGCGCATCGCACGCCGCCCGCGCGCGCGCGCGCGCGCATTCGAACGCTGCGCCCTATCAACT 240
DB 3963 GGGCGCATCGCACGCCGCCCGCGCGCGCGCGCGCGCGCATTCGAACGCTGCGCCCTATCAACT 4022

QY 241 TTCGATGTGTAGTCGCGTGCCTTACCATTGATGTGACCGCGGTGACGGGGAATCAGGGTTCGA 300
DB 4023 TTCGATGTGTAGTCGCGTGCCTTACCATTGATGTGACCGCGGTGACGGGGAATCAGGGTTCGA 4082

QY 301 TTCGGAGAGGAGGAGCTGAGAAACGGCTACCACATCCAAAGGAGGAGGAGGCGCGCAAA 360
DB 4083 TTCGGAGAGGAGGAGCTGAGAAACGGCTACCACATCCAAAGGAGGAGGAGGCGCGCAAA 4142

QY 361 TTACCCACTCCCG 373
DB 4143 TTACCCACTCCCG 4155

RESULT 9
US-10-151-081-17
; Sequence 17, Application US/10151081
; Publication No. US20030083293A1
; GENERAL INFORMATION:
; APPLICANT: Haglaczky, Gyula

```

```

; Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/151,081
; FILING DATE: 16-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/799,462
; FILING DATE: HERWITH 05-MAR-2001
; APPLICATION NUMBER: 09/724,693
; FILING DATE: HERWITH 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-151-081-17

Query Match 16.5%; Score 371.4; DB 14; Length 42999;
Best local Similarity 99.7%; Pred. No. 8.5e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCTCTCTCTCTGATACTGGTATTTCTAGAGCTAATACATGCCGACGGC 60
DB 3793 CTCGCTCTCTCTCTGATACTGGTATTTCTAGAGCTAATACATGCCGACGGC 3842
QY 61 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTTATCAGATCAAGACCAACCCCGTCAGC 120
DB 3843 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTTATCAGATCAAGACCAACCCCGTCAGC 3902
QY 121 CCCTCTCCGGCCCGCGGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 180
DB 3903 CCCTCTCCGGCCCGCGGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 3962
QY 181 GGGCGGATCGACGCCGCCCGCTGGCGGGGACGACCCATTCCGAACCTCTGCCCTATCAACT 240
|||||

; 3963 GGGCGGATCGACGCCGCCCGCTGGCGGGGACGACCCATTCCGAACCTCTGCCCTATCAACT 4022
; 241 TTCGATGTTAGTCCCGCTGCTTACCATGGTGACACACGGGTGACGGGAATCAGGGTTTCCA 300
; 4023 TTCGATGTTAGTCCCGCTGCTTACCATGGTGACACACGGGTGACGGGAATCAGGGTTTCCA 4082
; 301 TTCGGGAGAGGAGCGCTGAGAAACGGGTACACATCCAAAGGAGGACGAGCGCGCAAA 360
; 4083 TTCGGGAGGAGGAGCGCTGAGAAACGGGTACACATCCAAAGGAGGAGGACGAGCGCGCAAA 4142
; 361 TTACCCACTCCCG 373
; 4143 TTACCCACTCCCG 4155
;
; RESULT 10
; US-10-287-313-17
; Sequence 17, Application US/10287313
; Publication No. US20030101480A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/287,313
; FILING DATE: 01-No. US20030101480A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/724,726
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6869-402N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-287-313-17
```



```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M10098
; NAME/KEY: unsure
; LOCATION: (1)..(1969)
; OTHER INFORMATION: n - a or c or g or t
US-09-980-107-2231

Query Match      16.3%; Score 368.2; DB 10; Length 1969;
Best Local Similarity 99.2%; Pred. No. 8.5e-87;
Matches 370; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CTCGCTCCTCTCTCTACTTGGATAACTGTGGTAATCTCTAGAGCTAATACATCCGACGGGC 60
      |||
Db      228 CTCGCTCCTCTCTCTACTTGGATAACTGTGGTAATCTCTAGAGCTAATACATCCGACGGGC 287
      |||

QY      61 GCTGACCCCTTCGCGGGGGGATGCGTCATTTATCAGATCAAGACCAACCCGTCAGC 120
      |||
Db      288 GCTGACCCCTTCGCGGGGGGATGCGTCATTTATCAGATCAAAACCAACCCGTCAGC 347
      |||

QY      121 CCCTCTCCGGCCCCGGGGGGCGCGCGGCTTTGGTGACTCTAGATAACCTC 180
      |||
Db      348 CCCTCTCCGGCCCCGGGGGGCGCGCGGCTTTGGTGACTCTAGATAACCTC 407
      |||

QY      181 GGGCCGATGCACGCCCCCGTGCGGCGACGACCCATTGCGAAGCTCTGCCCTATCAACT 240
      |||
Db      408 GGGCCGATGCACGCCCCCGTGCGGCGACGACCCATTGCGAAGCTCTGCCCTATCAACT 467
      |||

QY      241 TTCGATGGTAGTCCCGCTGCTACCTACCTGGTGACACACGGGTACGGGGGAATCAGGGTTGCGA 300
      |||
Db      468 TTCGATGGTAGTCCCGCTGCTACCTACCTGGTGACACACGGGTACGGGGGAATCAGGGTTGCGA 527
      |||

QY      301 TTCGGGAGGGAGCGCTGAGAAACGGGTACACACATCCAAGGAAGCGACGCGCGCAAA 360
      |||
Db      528 TTCGGGAGGGAGCGCTGAGAAACGGGTACACACATCCAAGGAAGCGACGCGCGCAAA 587
      |||

QY      361 TTACCCACTCCCG 373
      |||
Db      588 TTACCCACTCCCG 600
      |||

RESULT 13
US-09-918-995-19104
; Sequence 19104, Application us/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918.995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19104
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19104

Query Match      16.0%; Score 360; DB 11; Length 427;
Best Local Similarity 100.0%; Pred. No. 4.9e-85;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1895 GGACGTGATAGCTCTGCTATTGTCAGGACCAATGATGGCTATCTTAAACGCTTAAGGAAAA 1954
      |||
Db      40 GGACGTGATAGCTCTGCTATTGTCAGGACCAATGATGGCTATCTTAAACGCTTAAGGAAAA 99
      |||

QY      1955 AACAAACACAGAACTCTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTT 2014
      |||

```

```
Db      18 TTACAC 13
|||||
RESULT 15
US-09-918-995-36028
; Sequence 36028, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36028
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36028

Query Match      15.5%; Score 349.8; DB 11; Length 393;
Best Local Similarity 93.1%; Pred. No. 2.3e-82;
Matches 366; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY      1542 GTATGTTCTGGCGGAGGGCCATCTGCATGGAACTTCTCACCACAGGCGCTGGAGCAG 1601
         |||||||
Db      1  GTATGTTCTGGTCGGAGGGGTCTCTGCATGGAAGTTATCACCACATGGCTGGAGCAG 60

QY      1602 TGCCTACTCCATAGATCAGTGCATGATCATGCAGATCAGTGCACACTGGTGAAGGGGAAAGC 1661
         |||||||
Db      61  TGCCTACTCCATATAGTCAGAGATCATGCTATCACTGCTCCTGCTGCTGGAAGGGGAAAGC 120

QY      1662 ACGAGTGCAGTTTGGAGGCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGCTCCTA 1721
         |||||||
Db      121  ACGAGTGCCTGTTGGAGGCCAACAAATGTCATACAGTCTGACAGAGCAGCAGCTCCTA 180

QY      1722 CAAGTCTTGGTGCAGATCCACGAAAAAAGCGGTGGTACACACCCCAAGAGACGG 1781
         |||||||
Db      181  CAAGTACTTGGTGCAATCCACGAAAAAAGCGGTGTTACACACCCCAAGAGACAG 240

QY      1782 CTAACCCCTGGAGTATCACCTTCCTCCCTCCAGGACCAGCTGACCAATACCTTTGA 1841
         |||||||
Db      241  TTAACCCCTGTAGTATGAGCTTACTTCCCTCCAGGACCAGCTGACCAATACCTTTGA 300

QY      1842 ATGCTGTATTTGGATCTCAGCTGCCTCTGTGGTTCCCTCCTCATTTTCTGGACGTG 1901
         |||||||
Db      301  ATGCTGGATTGGATCTCAGCTGCCTCTGTGGATCCCTCCTCATTTTCTGGACGTG 360

QY      1902 ATAGCTCTGCCATTGACAGGACAATGATGGCTA 1934
         |||||||
Db      361  ATAGCTCTGCCATTGACAGGACAATGATGGCTA 393
```

Search completed: August 4, 2003, 21:38:08  
Job time : 655 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:59:31 ; Search time 5661 Seconds

(without alignments)

9677.140 Million cell updates/sec

Title: US-10-005-549-1

Perfect score: 2254

Sequence: 1 ctcgctctctctctcttg.....tttgaaaaaaaaaaaaa 2254

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 4562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estmu:\*  
4: em\_estov:\*  
5: em\_estpl:\*  
6: em\_estro:\*  
7: em\_htc:\*  
8: em\_estl:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_fod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1217.4	54.0	1408	11 AK009324	AK009324 Mus muscu
2	1083.2	48.1	1201	9 AL540967	AL540967 AL540967
3	1026.6	45.5	1093	13 BU508098	BU508098 AGENCOURT
4	1025.2	45.5	1201	9 AL540968	AL540968 AL540968

C	5	945.6	42.0	1182	9	AL565939	AL565939
C	6	943	41.8	1095	13	BX404169	BX404169
C	7	912.4	40.5	1201	9	AL577525	AL577525
C	8	911.4	40.4	919	13	BX324562	BX324562
	9	899.8	39.9	1181	9	AL535478	AL535478
	10	893.4	39.6	1070	12	BM452095	BM452095
	11	885.8	39.3	1095	13	BX378721	BX378721
	12	885.4	38.4	895	14	CD107494	CD107494
	13	841.4	37.3	881	13	BQ212893	BQ212893
C	14	841.4	37.3	1201	9	AL526276	AL526276
C	15	840.2	37.3	1201	9	AL541381	AL541381
C	16	839.6	37.2	1201	9	AL548646	AL548646
C	17	834	37.0	1189	9	AL568878	AL568878
C	18	821.4	36.4	913	13	BU509450	BU509450
C	19	819	36.3	895	13	BX419858	BX419858
C	20	802.2	35.6	1201	9	AL573125	AL573125
C	21	800.2	35.5	1011	12	BM803766	BM803766
C	22	797.4	35.4	852	13	BQ228412	BQ228412
C	23	794	35.2	808	11	AF116721	AF116721
C	24	794	35.2	1238	13	BU515543	BU515543
C	25	789.6	35.0	1201	9	AL554925	AL554925
C	26	775.6	34.4	874	13	BU166850	BU166850
C	27	770.4	34.2	873	13	BQ721866	BQ721866
C	28	752.4	34.1	949	13	BQ958149	BQ958149
C	29	747	33.1	779	12	BQ014708	BQ014708
C	30	733	32.5	735	13	BU625614	BU625614
C	31	728.4	32.3	1100	13	BQ278712	BQ278712
C	32	717	31.8	922	13	BQ218761	BQ218761
C	33	714.8	31.7	1201	9	AL526309	AL526309
C	34	714.6	31.6	822	12	BM981520	BM981520
C	35	712.6	31.6	916	13	BM981520	BM981520
C	36	705	31.3	837	12	BM011995	BM011995
C	37	703.8	31.2	900	10	BE546982	BE546982
C	38	703.4	31.2	711	12	BM715207	BM715207
C	39	703.4	31.2	973	10	BQ029803	BQ029803
C	40	702.6	31.2	973	10	BQ029803	BQ029803
C	41	697	30.9	755	14	CA453717	CA453717
C	42	691	30.7	743	9	AI207712	AI207712
C	43	686.2	30.4	875	14	CD243557	CD243557
C	44	680.8	30.2	864	12	BI655207	BI655207
C	45	680.8	30.2	864	12	BI655207	BI655207

#### ALIGNMENTS

RESULT 1	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
LOCUS	AK009324				
DEFINITION	AK009324				
ACCESSION	AK009324				
VERSION	AK009324.1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1				
AUTHORS	Carninci,P. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,N., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				







1615 GAGTCACTGATGATGAGTCACTGAGTCCACACTGGTGAAGGGGAAAGCAGCAGTGCAGTTT 1674  
1675 GGAGCCCAACAACTCTCAATACAGTCTCACAAGACAGCAGTCCCTACAGTCTCTGGTG 1734  
1735 CAGATCCACGAAAAACGGCTGTGTACACACACCCCAAGAGAGAGGCTAACCTCGAGT 1794  
479 CAGATCAGCAAGAAAAACGGCTGTGTACACACACCCCAAGAGAGAGGCTAACCTCGAGT 420  
1795 ATCACCCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1854  
419 ATCACCCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 360  
1855 ATCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1914  
359 ATCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
1915 TTGAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1974  
299 TTGAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
1975 CAAGTACTCAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2034  
239 CAAGTACTCAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
2035 GCATCTCTATTAAGAAACCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094  
179 GCATCTCTATTAAGAAACCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
2095 AGACTAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2154  
119 AGACTAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
2155 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2210  
59 MAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4

RESULT 3  
BU508098  
LOCUS BU508098 1093 bp mRNA linear EST 12-SEP-2002  
DEFINITION AGENCOURT\_10128262 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:6502128  
5', mRNA sequence.  
ACCESSION BU508098  
VERSION BU508098.1 GI:22814331  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1093)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L14M14057 row: 1 column: 01  
High quality sequence stop: 869.  
Location/Qualifiers  
1. 1093  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/clone="IMAGE:6502128"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_71"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2.1 kb."  
BASE COUNT 326 a 274 c 251 g 239 t 3 others  
ORIGIN  
Query Match 45.5%; Score 1026.6; DB 13; Length 1093;  
Best Local Similarity 97.5%; Pred. No. 2.3e-152;  
Matches 1052; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
QY 1026 GCTGGATCAACCCCTGCCAGCAGCAGTGCACACAGGAGAGCGTCTCTTCAGAGATGA 1085  
DB 11 GCTGGATCAACCCCTGCCAGCAGCAGTGCACACAGGAGAGCGTCTCTTCAGAGATGA 70  
QY 1086 AGATGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGAGGA 1145  
DB 71 AGATGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGAGGA 130  
QY 1146 GCAGCTGAGGAGCAAGAAATCTGAAGATGATGCGATGGAAAAGAAACTTGGCCATCCT 1205  
DB 131 GCAGCTGAGGAGCAAGAAATCTGAAGATGATGCGATGGAAAAGAAACTTGGCCATCCT 190  
QY 1206 AGAGAAAATTAAGAAACACAGAGCAAGATTAATTAATGCTGAGTGTGCGCTCGGT 1265  
DB 191 AGAGAAAATTAAGAAACACAGAGCAAGATTAATTAATGCTGAGTGTGCGCTCGGT 250  
QY 1266 GCAGGCACTGACCGCTGATGAAGGAGCTGAGGATATATACCGATCACAGATTCMA 1325  
DB 251 GCAGGCACTGACCGCTGATGAAGGAGCTGAGGATATATACCGATCACAGATTCMA 310  
QY 1326 AGCGGAAACTATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1385  
DB 311 AGCGGAAACTATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 370  
QY 1386 CCTCAAGTTGACGAGCAGCGCTTGCACACAGATCTCCAGATCTCCAGATCTCCAGATCT 1445  
DB 371 CCTCAAGTTGACGAGCAGCGCTTGCACACAGATCTCCAGATCTCCAGATCTCCAGAT 430  
QY 1446 AGAGCGCACTTCATCTTAACTTTTCCCTTTAAAGATAAATTTCCCTTTGACCCACC 1505  
DB 431 AGAGCGCACTTCATCTTAACTTTTCCCTTTAAAGATAAATTTCCCTTTGACCCACC 490  
QY 1506 ATTTGTCAGGGTTGTGCTCCAGTCTCTGAGGGTATGTTCTGGGGGAGGGGCCAT 1565  
DB 491 ATTTGTCAGGGTTGTGCTCCAGTCTCTGAGGGTATGTTCTGGGGGAGGGGCCAT 550  
QY 1566 CTGCATGGAACTTCTCACCACAGCGCTGGAGCAGTGCCTACTCTCATAGAGTCAGTGT 1625  
DB 551 CTGCATGGAACTTCTCACCACAGCGCTGGAGCAGTGCCTACTCTCATAGAGTCAGTGT 610  
QY 1626 CATGCAGATCAGTGCACACTGCTGAGGGGAAAGCAGCAGTGCAGTGTGAGGCCAACAA 1685  
DB 611 CATGCAGATCAGTGCACACTGCTGAGGGGAAAGCAGCAGTGCAGTGTGAGGCCAACAA 670  
QY 1686 ATCTCAATACAGTCTCACAAGACACAGCAGTCTCACAAGTCTCTGGTGCAGATCCACGA 1745  
DB 671 ATCTCAATACAGTCTCACAAGACACAGCAGTCTCACAAGTCTCTGGTGCAGATCCACGA 730  
QY 1746 AAAAAAGGCTGTGTACACACCCCAAGAGAGCGCTAACCTCGAGTATACCCCTTCC 1805  
DB 731 AAAAAAGGCTGTGTACACACCCCAAGAGAGCGCTAACCTCGAGTATACCCCTTCC 790  
QY 1806 TCCTCCCGGAGCAGCAGTGCAGCAATTAACCTTTGAATGCTGATTTGGATCTCAGCTG 1865  
DB 791 TCCTCCCGGAGCAGCAGTGCAGCAATTAACCTTTGAATGCTGATTTGGATCTCAGCTG 850  
QY 1866 CCTCTGTGGTTCCCTCCCTCATTTTTCCTGGAGCTGATAGCTGCTGCTATTTGAGGACAA 1925

Db 851 CTTCTGTGGTTCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCCTATTGCCANNAGCA 910  
 QY 1926 TGATGGCTATTCTAAACGCTTAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAA 1985  
 Db 911 TGATGGCTATTCTAAACGCTTAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAA 970  
 QY 1986 GACTGACATTACAGACCAACACCACTTGTGGAACTCTGCTAGCAGGCTTCTTTATA 2045  
 Db 971 GACTGACATTACAGACCAACCACTTGTGGAACTCTGCTAGCAGGCTTCTTTATA 1030  
 QY 2046 AAGAACTCTGAGCCTCTTATATGCTGGAACCTGCTGCTCCAGCTAGAGC 2104  
 Db 1031 AAAAAAATCTCCAGCCTCTTATA-TGCTGGAACTCAGCTGGGCTCAAACTAAACC 1088

## RESULT 4

AL540968

LOCUS

DEFINITION AL540968 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE004Y104 5-PRIME. mRNA sequence.

ACCESSION

AL540968

VERSION

AL540968.2

GI:30544685

EST.

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished

On Feb 15, 2001 this sequence version replaced gi:12871590.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 78.r For more

information about this cluster, see http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DE004BE02Q1&amp;cluster=78.r. Contact :

Feng Liang Email: fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DE004BE02Q1.

Location/Qualifiers

1. .1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DE004Y104"

/tissue\_type="PLACENTA"

/clone\_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with NotI and cloned into

the NotI and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 273 a 303 c 400 g 201 t 24 others

ORIGIN

Query Match 45.5%; Score 1025.2; DB 9; Length 1201;

Best Local Similarity 95.8%; Pred. No. 3.7e-152;

Matches 1084; Conservative 18; Mismatches 23; Indels 7; Gaps 5;

QY 365 CCACATCCCGGAGGTGGCGGCGGCGGCTTGGCGAAGGGGATCAGGAAGTGGCGGACC 424

Db 71 CTTGTCGGGAGGTGGCGGCGGCGGCGGCTTGGCGAAGGGGATCAGGAAGTGGCGGACC 130

QY 425 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 484

Db 131 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 187

QY 485 CCCGCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544

Db 188 CCCCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 247  
 QY 545 AGCAGCCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 604  
 Db 248 AGCAGCCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 307  
 QY 605 GCCCAGGG 664  
 Db 308 GCCCAGGG 367  
 QY 665 CCATCTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 724  
 Db 368 YCATCTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427  
 QY 725 GCTCGGAGTTCTCTGCTGGCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 784  
 Db 428 GCTCGGAGTTCTCTGCTGGCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487  
 QY 785 ATCTCCCCCAGCGGGGCTGGTGGGATCTGCTGGGATCTGCTGGGATCTGCTGGGATCTGCT 844  
 Db 488 ATCTCCCCCAGCGGGGCTGGTGGGATCTGCTGGGATCTGCTGGGATCTGCTGGGATCTGCT 546  
 QY 845 AGTCATACCTGCTGTCGCCCCCATCTGGTGGGAGTCTGATGACCTAACCTGCTGCTG 904  
 Db 547 AGTCATACCTGCTGTCGCCCCCATCTGGTGGGAGTCTGATGACCTAACCTGCTGCTGCTG 606  
 QY 905 CTGCTTTGGAGAGCTGGTGGACATAAAGAAAGGAATACTCTGCTTATTTGACGATCTGA 964  
 Db 607 CTGCTTTGGAGAGCTGGTGGACATAAAGAAAGGAATACTCTGCTTATTTGACGATCTGA 666  
 QY 965 AGAGGATCATCTCCGAGCTGTGTAACCTCTATAACCTCTCCCTCAGCATCCAGATGTGAGA 1024  
 Db 667 AGAGGATCATCTCCGAGCTGTGTAACCTCTATAACCTCTCCCTCAGCATCCAGATGTGAGA 726  
 QY 1025 TGCTGGATCAACCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1084  
 Db 727 TGCTGGATCAACCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 786  
 QY 1085 AAGATGAGGAGATGCTTGGAGACACAGAACTTAGATCACTATGAATGAAGAGGAGGAG 1144  
 Db 787 AAGATGAGGAGATGCTTGGAGACACAGAACTTAGATCACTATGAATGAAGAGGAGGAG 846  
 QY 1145 AGCCAGCTGAGGGCAAGAAATCTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 1204  
 Db 847 AGCCAGCTGAGGGCAAGAAATCTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 906  
 QY 1205 TAGAGAAATTTAAAGAACAGAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1264  
 Db 907 TAGAGAAATTTAAAGAACAGAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 966  
 QY 1265 TGCAGGCGCTGACCGGCTGATGAAGAGGCTCAGGGATATATACCGATCAGAGTTTCA 1324  
 Db 967 TGCAGGCGCTGACCGGCTGATGAAGAGGCTCAGGGATATATACCGATCAGAGTTTCA 1026  
 QY 1325 AAGCGGAACTATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1384  
 Db 1027 AAGCGGAACTATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1085  
 QY 1385 TCCTCAAGTTGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1444  
 Db 1086 TCCTCAAGTTGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1145  
 QY 1445 AAGCGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1496  
 Db 1146 AAG--GGCGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1195

RESULT 5

AL565939/c

LOCUS

DEFINITION

AL565939 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone

CS0DF009Y108 3-PRIME, mRNA sequence.

ACCESSION

AL565939

AL565939

1182 bp

mrna

linear

EST 12-MAY-2003



Faraday Avenue Genoscope sequence ID : CLOBB0172H1RP1.					
FEATURES	Location/Qualifiers				
source	1..1095				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CLOBB0172H11"				
	/tissue_type="NEUROBLASTOMA"				
	/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dP) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."				
BASE COUNT	260 a 293 c 350 g 162 t	30 others			
ORIGIN					
Query Match	41.88;	Score 943;	DB 13;	Length 1095;	
Best Local Similarity	94.7%;	Pred. No. 3.6e-139;			
Matches 990;	Conservative 24;	Mismatches 24;	Indels 7;	Gaps 4;	
QY	420	GGACCGGCGGGCGGGCGGCGGCGGGCGGGAGC--CCGAGCGCAGGCCGGAGGC	478		
Db	54	GGATCGGGCGGGCGGGCGGGCGGGCGGGCGGGAGCAGGAGCGCAGCGCGGAGCG	113		
QY	479	TCCGGGCCCCCGGGCCCCGGAGGGAGCG---GAGCGGAGGATGCAAGCACCCAGCCGACGCCG	535		
Db	114	TCCGGGCCCCCGGGCCCCGGAGGGAGCGCACGCGCCAAAGGATGCAGCAGCCGACGCCG	173		
QY	536	AGGGGACGACGAGCGCGGGCGGGCGGAGCTGGGGGCCAGGGGGCGGCCCGGGGGG	595		
Db	174	AGGGGACGACGAGCGGGGGCGGGGCGAGCGCGGGGGGGCCAGGGGGCGGCCGGGGG	233		
QY	596	CCGGGGCGGCCACGAGGGGGGGGGCGGGGGCGGGGCCCTGCTGAGCGAGAGCTGAAGC	655		
Db	234	CCGGGGCGGCCACGAGGGGGGGGGCGGGGGCGGGGCCCTGCTGAGCGAACAAGAAC	293		
QY	656	TGCTGAGTCCATCTTCACCGGGGCCACAGAGGCTTCGCCANTTGCCAGGCGCTGCGCTGG	715		
Db	294	TGCTGAGTCCATCTTCACCGGGGCCACARAGCTTCGCGATTGGCCAGSCCTGCGCTGG	353		
QY	716	ACGAGCTGAGCTGCGAGTTCTGTGCTGGGCGCGGAGGGCGGGGGGGGGCGCGCGC	775		
Db	354	ACRARTGACWGRATWCCTGCTGCTGGGCGCGGARGGGCGGGGGCGGGCGCGCGC	413		
QY	776	CCGGACCGCATFCCCACCAGGGGGTCTGGTCTGGGGATCTCTGTCGCCGATCCACTGCA	835		
Db	414	CCGGACCGCATCTCCCCCACCAGGGGCTCGGTCTGGGGATCTCTGTCGCCGATCCACTGCA	473		
QY	836	ACATCAGGAGTCATACCCCTGCTGGCCCCCCCCATCTGGTCGGTGGAGTCTGATGACCCCTA	895		
Db	474	ACATCAGGAGTCATACCCCTGCTGGCCCCCCCCATCTGGTCGGTGGAGTCTGATGACCCCTA	533		
QY	896	ACTTGGCTGCTGTCTTGGAGAGGCTGTGGACATAAAGAAGGAATACCTCTGCTATTGC	955		
Db	534	ACTTGGCTGCTGTCTTGGAGAGGCTGTGGACATAAAGAAGGAATACCTCTGCTATTGC	593		
QY	956	AGCATCTGAAGAGGATCATCTCCGACCTGTGTAAACTCTATACTCCCTCAGCATCCAG	1015		
Db	594	AGCATCTGAAGAGGATCATCTCCGACCTGTGTAAACTCTATACTCCCTCAGCATCCAG	653		
QY	1016	ATGTGGAGATGCTGGATCAACCTTTGCCAGCAGCAGTAGTCACACAGGAAGAGCTGTCTT	1075		
Db	654	ATGTGGAGATGCTGGATCAACCTTTGCCAGCAGCAGTAGTCACACAGGAAGAGCTGTCTT	713		
QY	1076	CAGAAGATGAAGATGAGGAGATCCCTGAGGACACAGAAGACTTAGATCACTATGAATGA	1135		
Db	714	CAGAAGATGAAGATGAGGAGATCCCTGAGGACACAGAAGACTTAGATCACTATGAATGA	773		
QY	1136	AAGAGGAGAGCCAGCTGAGGGCAGAAAATCTGAGATGATGGCATTTGGAAAAAGAAACT	1195		
Db	774	AAGAGGAGAGCCAGCTGAGGGCAGAAAATCTGAGATGATGGCATTTGGAAAAAGAAACT	833		

```
Db 975 CAAGATCATTAAATGKAGAGTCTGGTCGGTCAGGCCCACTGACCGCTGATGAAG 916
Qy 1291 GAGCTCAGGATATATACCGATCACAGAGTTTCAAGCGGGAAGAACTATGAGTCGAAGTC 1350
Db 915 GRCCTCAGGATATATACCGATCACAGAGTTTCAAGCGGGAAGAACTAGGAGTCGAAGTC 856
Qy 1351 GTGAATGACAGCTGATGATGGAATGTCAAACTCTCAAGTTGACGAGGACGGCT 1410
Db 855 GTGAATACAGCTGATGATGGAATGTCAAACTCTCAAGTTGACGAGGACGGCT 796
Qy 1411 TTGCACACAGATCTCCAGATCTCAAGAGAGAAAGAGCGGCACTTCACTACTTAAC 1470
Db 795 TTGCACACAGATCTCCAGATCTCAAGAGAGAAAGAGCGGCACTTCACTACTTAAC 736
Qy 1471 TTTTCCCTTAAAGATAACTTTCCCTTTGACCCACCACTTTGTGAGGGTTGTCTCCAGTC 1530
Db 735 TTTTCCCTTAAAGATAACTTTCCCTTTGACCCACCACTTTGTGAGGGTTGTCTCCAGTC 676
Qy 1531 CTCTCTGGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAACCTTCTACCAACAG 1590
Db 675 CTCTCTGGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAACCTTCTACCAACAG 616
Qy 1591 GSCCTGGAGCAGTCTACTCTCATAGATGATGATGATGATGATGATGATGATGATGATG 1650
Db 615 GSCCTGGAGCAGTCTACTCTCATAGATGATGATGATGATGATGATGATGATGATGATG 556
Qy 1651 AAGGGGAAGCAGCAGTGTGAGTGTGGAGCCCAACAAATCTCAATACAGTCTGACAAAGCA 1710
Db 555 AAGGGGAAGCAGCAGTGTGAGTGTGGAGCCCAACAAATCTCAATACAGTCTGACAAAGCA 496
Qy 1711 CAGCAGTCTCAAGTCTCTGGTGAGATCCACGAAACAAACAGGCTGGTACACACCCCA 1770
Db 495 CAGCAGTCTCAAGTCTCTGGTGAGATCCACGAAACAAACAGGCTGGTACACACCCCA 436
Qy 1771 AAGAGAGCAGCTAACCTGAGTATCACCTCTCCCTCCAGGACCACTGGACCA 1830
Db 435 AAGAGAGCAGCTAACCTGAGTATCACCTCTCCCTCCAGGACCACTGGACCA 376
Qy 1831 ATTACCTTGAATGCTGATTTGGTATCTCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 1890
Db 375 ATTACCTTGAATGCTGATTTGGTATCTCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 316
Qy 1891 TCCTGGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1949
Db 315 TCCTGGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
Qy 1950 AAAAAACAAACACAGACGTTTCAAGTACTCAAGACTGACCTTACAGACCAACCAACC 2009
Db 255 AAAAAACAAACACAGACGTTTCAAGTACTCAAGACTGACCTTACAGACCAACCAACC 196
Qy 2010 ACCTTGTGGAACTTCTGCTAGAGGCACTTCTTATAAAGAACTTTTCGAGCCTCCTTAT 2069
Db 195 ACCINGCTGGAACCTTCTGCTAGAGGCACTTCTTATAAAGAACTTTTCGAGCCTCCTTAT 136
Qy 2070 ATTGCTGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2129
Db 135 ATNGCTGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 76
Qy 2130 AATTATTTTCTCTTATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 2166
Db 75 AATTATTTTCTCTTATTTAATGATGATGATGATGATGATGATGATGATGATGATGATG 39
```

## RESULT 8

```
LOCUS BX324562/c
DEFINITION BX324562 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DC025YA19 3-PRIME, mRNA sequence.
ACCESSION BX324562
VERSION BX324562.1 GI:30309380
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 919)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 9191006 EVRY cedex - France
Email: seq@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 78.r For more
information about this cluster, see http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AC025AA10NP1&cluster=78.r. Contact:
Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS0AC025AA10NP1.
FEATURES
Location/Qualifiers
1..919
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC025YA19"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 239 a 185 c 229 g 255 t 11 others
ORIGIN
Query Match 40.4%; Score 911.4; DB 13; Length 919;
Best Local Similarity 98.8%; Pred. No. 3.7e-134;
Matches 906; Conservative 10; Mismatches 1; Indels 0; Gaps 0;
Qy 1304 TATACCGATCACAGAGTTTCAAGCGGGAAGAACTATGATGATGATGATGATGATGATG 1363
Db 919 TATACCGATCACAGAGTTTCAAGCGGGAAGAACTATGATGATGATGATGATGATGATG 860
Qy 1364 TGATGATGGAATGTCAAACTCTCTCAAGTTGACGAGGACGCTTTGACCAAGATC 1423
Db 859 TGATGATGGAATGTCAAACTCTCTCAAGTTGACGAGGACGCTTTGACCAAGATC 800
Qy 1424 TCAGATCTCTCAAGAGAAAGAGGAGCGGCACTTCTTACTTAACTTTTCTTTAAAG 1483
Db 799 TCAGATCTCTCAAGAGAAAGAGGAGCGGCACTTCTTACTTAACTTTTCTTTAAAG 740
Qy 1484 ATAACTTCTCTTTGACCCACCACTTGTGTCAGGGTTGTGTCCTGAGGAGGT 1543
Db 739 ATAACTTCTCTTTGACCCACCACTTGTGTCAGGGTTGTGTCCTGAGGAGGT 680
Qy 1544 ATGTTCTGGCGGAGGGCCATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1603
Db 679 ATGTTCTGGCGGAGGGCCATCTGATGATGATGATGATGATGATGATGATGATGATGATG 620
Qy 1604 CCTACTCCATAGAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1663
Db 619 CCTACTCCATAGAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 560
Qy 1664 GAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTCTTACA 1723
Db 559 GAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCTCTTACA 500
Qy 1724 AGTCTTGTGTCAGATCCAGCAAAAAAAGCGCTGTGTACACACCCCAAGAGAGAGCGCT 1783
Db 499 AGTCTTGTGTCAGATCCAGCAAAAAAAGCGCTGTGTACACACCCCAAGAGAGAGCGCT 440
Qy 1784 AACCTTGGAGTATCACCTTCTCTCCCTCCAGGACCACTGGACCAATTTACCTTTGAAT 1843
Db 439 AACCTTGGAGTATCACCTTCTCTCCCTCCAGGACCACTGGACCAATTTACCTTTGAAT 380
```









```
Query Match      38.4%; Score 865.4; DB 14; Length 895;
Best Local Similarity 98.7%; Pred. No. 6.8e-127;
Matches 883; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
```



|||||  
 543 AACATGACGAGTGAATGACAGCTCTGTATGATGAATGCTCAACTCTCTCAA 502  
 1393 GTTGACGAGGAGCGCTTTGACACAGATCTCCAGATCTCTCAAGAGAGAGAGAGCC 1452  
 603 GTTGACGAGGAGCGCTTTGACACAGATCTCCAGATCTCTCAAGAGAGAGAGAGCC 662  
 1453 GACTTCATCTACTTAACTTTTCCCTTTAAAGATACTTCCCTTTGACCCACCATTTGTC 1512  
 663 GACTTCATCTACTTAACTTTTCCCTTTAAAGATACTTCCCTTTGACCCACCATTTGTC 722  
 1513 AGGGTTGTCTCTCCAGTCTCTCTGGAGGATGTTCTGGGGGAGGGGCCATCTGCATG 1572  
 723 AGGGTTGTCTCTCCAGTCTCTCTGGAGGATGTTCTGGGGGAGGGGCCATCTGCATG 782  
 1573 GAATCTCTCAACAAAGAGGCTGGAGCAGTCTCTACTCATAGATCAGTATGATGAG 1632  
 783 GAATCTCTCA-CAACAGAGGCTGGAGCAGTCTCTACTCATANAGTCTAGTATGATGAG 841  
 1633 ATCAGTCCACACTGGT--GAAGGGGAAGACAGAGTGC 1669  
 842 ATCAGTCCCACTGGTGAAGAGGGAAGACAGAGTGC 880

## RESULT 14

## AL526276/c

LOCUS 1201 bp mRNA linear EST 23-MAY-2003  
 DEFINITION AL526276 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
 CDNA clone CS0DC016Yp18 3-PRIME, mRNA sequence.

## ACCESSION

AL526276

## VERSION

AL526276.2 GI:31064137

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

On Feb 13, 2001 this sequence version replaced gi:12789769.

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 78.r For more

information about this cluster, see http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DC016DH09NP1&amp;cluster=78.r. Contact :

Feng Liang Email : fliang@life.techn.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DC016DH09NP1.

## FEATURES

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DC016Yp18"

/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 290 a 200 c 263 g 321 t 127 others

## ORIGIN

Query Match 37.3%; Score 841.4; DB 9; Length 1201;

Best Local Similarity 82.4%; Pred. No. 3.8e-123;

Matches 914; Conservative 77; Mismatches 109; Indels 9; Gaps 4;

Db 1103 AARPRRARSAAAAAGRAAGTTTTAVGAGTSRGTSSRKYCTYGGMAAADAHTTTSTT 1044  
 QY 1139 AGGAAGACCGAGCTGAGGGCAAGAAATCTGAAGATGATGGCATTTGGAAGAAACATTTGG 1198  
 Db 1043 ATTTAAATVAASSSSVSCMBCTSGSCAAVAATHVTASSTSHBSATTSMAAAMAAAVTT 984  
 QY 1199 CCATCCTAGAGAAATTAATAAGAACACAGAGGCAAGATTACTTAATAGGTGCGAGTGTCTG 1258  
 Db 983 CCMATCTAWVDAAATTAATAAARAAACAGVCCCAATTAATTAATSYNCAST-----TTTT 930  
 QY 1259 GCTCGTGCAGCGCACTGACCGGCTGATGAAGGAGCTCAGGAGATATATACGATCACACAGA 1318  
 Db 929 SLYCGTAKMSMAACTSNCCSSCTSWTGMSSSSCTCCSSSATWATCCCTCTCMCCRG 870  
 QY 1319 GTTCAAAGGGCGGAACCTATGACGTCGAACTCGTGAATGACAGTCTGTATGATTTGGAATG 1378  
 Db 869 AGTTTAAAGCGGGAACCTATGACGYSGAACCTGTAATGACAGTCTGTATGATTTGGAATG 810  
 QY 1379 TCAAACT--CCTCAAAGTTGAC--CAGGACAGCGCTTTGCAACAGCATCTCCAGATCTCAA 1436  
 Db 809 TCAAACTCCCTCAAGTTGACMCAGGACAGCGCTTTGCAACAGCATCTCCAGATCTCAA 750  
 QY 1437 AGAGAAAGAGGCGGACTTCACTCTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTT 1496  
 Db 749 AGAGAAAGAGGCGGACTTCACTCTACTTAACTTTTCCCTTTAAAGATAACTTTCCCTT 690  
 QY 1497 TGACCCACCATTTGTGAGGGTTGTCTCTCCAGTCTCTCTCGAGGGTATGTTCTGGGGGG 1556  
 Db 689 TGACCCACCATTTGTGAGGGTTGTCTCTCCAGTCTCTCTCGAGGGTATGTTCTGGGGGG 630  
 QY 1557 AGGGGCCATCTGCATGGAACCTTCTACAAACAGAGGCTGGGGGAGTGCCTACTCCATAGA 1616  
 Db 629 AGGGGCCATCTGCATGGAACCTTCTACAAACAGAGGCTGGGGGAGTGCCTACTCCATAGA 570  
 QY 1617 GTCAGTGATCATGAGATCAGTGCACACTGGTGAAGGGGAAACACAGAGTGCAGTTGG 1676  
 Db 569 GTCAGTGATCATGAGATCAGTGCACACTGGTGAAGGGGAAACACAGAGTGCAGTTGG 510  
 QY 1677 AGCCAAACAAATCTCAATACAGTCTGACAAGAGCAGCAGTCTTCAAGTCTCTTGGTGCA 1736  
 Db 509 AGCCAAACAAATCTCAATACAGTCTGACAAGAGCAGCAGTCTTCAAGTCTCTTGGTGCA 450  
 QY 1737 GATCACGAAACAAACGGCTGGTACACACCCCAAGAGAGAGGCTTAACCTCGAGTAT 1796  
 Db 449 GATCACGAAACAAACGGCTGGTACACACCCCAAGAGAGAGGCTTAACCTCGAGTAT 390  
 QY 1797 CACCTTCTCTCCCTCCAGCAGCAGTGCACCACTTACCTTGAATCTCTATTTGGAT 1856  
 Db 389 CACCTTCTCTCCCTCCAGCAGCAGTGCACCACTTACCTTGAATCTCTATTTGGAT 330  
 QY 1857 CTCAGCTGCTCTCTGCTGCTCCCTCCCTCAATTTCTCTGGAGCTGATAGCTCTGCCATT 1916  
 Db 329 CTCAGCTGCTCTCTGCTGCTCCCTCCCTCAATTTCTCTGGAGCTGATAGTCTGCCATT 270  
 QY 1917 GCAGACAAATGATGGCTATTCTAAACGCTAAGGAGAGAGAGAGAGAGAGAGAGAGAG 1976  
 Db 269 GCAGACAAATGATGGCTATTCTAAACGCTAAGGAGAGAGAGAGAGAGAGAGAGAGAG 210  
 QY 1977 AGTACTCAAGTACTGACTTACAGACCAACACACCTTGGCTGGAGAGAGAGAGAGAGAG 2035  
 Db 209 AGTACTCAAGTACTGACTTACAGACCAACACACCTTGGCTGGAGAGAGAGAGAGAGAG 150  
 QY 2036 CATCTTTATAAAGAAACTTTTCGAGCTCTCTTATATGCTGGAACCTCAGCTGCTGCTCA 2095  
 Db 149 CATCTTTATAAAGARACTTTTCGAGCTCTCTTATATGCTGGAACCTCAGCTGCTGCTCA 90  
 QY 2096 GACTAGAGCTCTCTTACCTATGCTATGAGATTTTAAATTTATTTCTTTATTTTATGAT 2155  
 Db 89 GACTAGAGCTCTCTTACCTATGCTATGAGATTTTAAATTTATTTCTTTATTTTATGAT 30  
 QY 2156 ACTGCTTTTGGTTTACAGTGTATGATG 2184  
 Db 29 ACTGCTTTTGGTTTACAGTGTATGATG 1



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:35:23 ; Search time 83 Seconds  
(without alignments)  
807.019 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQPGQQPGGCGQLGG.....KSLVQIHKNWYTPPKEDG 422

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2229	100.0	422	AAE24493	Human RAT1d6 (reg
2	1863	83.6	374	AAE24493	Human colon cancer
3	1509	67.7	291	AAE24493	Human prostate can
4	1300.5	58.3	340	AAE24493	Novel human protei
5	1293	58.0	245	AAE24493	Human RAT1d6 tran
6	1266	56.8	263	AAE24493	Human ORFX ORF1537
7	1120.5	50.3	308	AAE24493	HTRM clone 3340290
8	1020.5	45.8	397	AAE24493	Drosophila melanog
9	834	37.4	164	AAE24493	Human RAT1d6 ubiq

10	808	36.2	217	22	ABB67021	Drosophila melanog
11	440.5	19.8	306	22	ABB60719	Drosophila melanog
12	440.5	19.8	306	22	ABB60722	Drosophila melanog
13	304	13.6	747	22	ABG11969	Novel human diagno
14	260	11.7	400	22	ABG28233	Novel human diagno
15	187.5	8.4	1909	22	ABG15534	Novel human diagno
16	163	7.3	452	22	ABG20796	Novel human diagno
17	154	6.9	59	22	ABG03538	Novel human diagno
18	140.5	6.3	46	22	ABG54524	Human liver peptid
19	140.5	6.3	46	22	ABG39508	Peptide #7014 enco
20	140.5	6.3	46	22	AAE60200	Human brain expres
21	140.5	6.3	46	22	AAE72820	Human bone marrow
22	140.5	6.3	46	22	AAE33050	Peptide #7087 enco
23	140.5	6.3	46	22	ABG42649	Human peptid enco
24	135.5	6.1	295	23	AAE28245	Yeast Cdc34 protei
25	135	6.1	199	22	ABG63741	Drosophila melanog
26	134.5	6.0	44	22	ABG03535	Novel human diagno
27	134.5	6.0	223	24	ABJ26294	Aspergillus fumiga
28	133	6.0	238	24	ABJ25694	Aspergillus fumiga
29	131	5.9	244	23	ABP73432	Candida albicans e
30	128.5	5.8	190	22	ABG63097	Drosophila melanog
31	128	5.7	226	23	ABP41187	Human ovarian anti
32	127.5	5.7	207	21	ABE43031	Human ORFX ORF2795
33	127.5	5.7	207	24	ABP71416	Human UBE2 related
34	125	5.6	81	22	AAO02768	Human polypeptide
35	125	5.6	1235	23	ABB77435	Human tumour marke
36	123	5.5	463	21	ABA41079	Human ORFX ORF843
37	123	5.5	914	23	ABP43569	FLJ10378 fls clone
38	123	5.5	963	24	ABJ19401	Human intracellular
39	122.5	5.5	177	18	AAW31278	Clam cyclin-select
40	122.5	5.5	1048	21	AAE42118	Arabidopsis thalia
41	122	5.5	477	22	ABE59034	Drosophila melanog
42	121	5.4	542	22	ABG20670	Novel human diagno
43	120.5	5.4	716	14	AAE30730	B. burgdorferi 79
44	120.5	5.4	924	23	ABE54116	Lactococcus lactis
45	120.5	5.4	1004	22	AAE31922	C glutamicum prote

#### ALIGNMENTS

##### RESULT 1

AAE24493  
ID AAE24493 standard; Protein; 422 AA.

XX AAE24493;

AC AAE24493 (first entry)

DT 04-OCR-2002 (first entry)

XX Human RAT1d6 (regulated in activated T-lymphocyte ld6) protein.

Human; ubiquitin conjugating enzyme; UBC; RAT1d6; immune disorder; regulated in activated T-lymphocyte ld6; neuronal disorder; cancer; tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma; sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma; multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia; depression; epilepsy; acquired immuno deficiency syndrome; allergy; AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological; myocardial infarction; renal tubular acidosis; gonadal dysgenesis; dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV; anticonvulsant; antiinflammatory; Cushing's syndrome; cardiant; ophtalmological.

XX Homo sapiens.

XX Key Location/Qualifiers  
XX Domain 89..333  
/note= "Transmembrane domain"

XX WO200236741-A2.

XX 10-MAY-2002.

XX

PF 29-OCT-2001; 2001WO-US46559.  
 XX  
 PR 30-OCT-2000; 2000US-244688P.  
 PR 30-JUL-2001; 2001US-308706P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Bowen MA, Wu Y, Yang W, Finger JN;  
 XX  
 DR WPI; 2002-479758/51.  
 DR N-PSDB; RAD39359.  
 XX  
 XX Novel ubiquitin conjugating enzyme polypeptide isolated from activated  
 PT human T cell, for screening modulators useful for treating cancer,  
 PT immune disorder, lymphoproliferative disorder, neurodegenerative  
 PT disorder -  
 XX  
 PS Claim 6; Fig 3; 169pp; English.  
 XX  
 CC The invention relates to a novel ubiquitin conjugating enzyme (UBC)  
 CC homologue, RAT1d6 (regulated in activated T-lymphocytes 1d6) and its  
 CC corresponding nucleic acid. The invention also relates to methods for  
 CC treating, diagnosing, preventing and screening for disorders related  
 CC to the expression of RAT1d6. UBC is useful for screening for candidate  
 CC compounds capable of binding to and/or modulating its activity. UBC is  
 CC useful for treating an immune or neuronal disorder in a mammal. The  
 CC method is useful for treating a cancer or tumour. It is useful for  
 CC suppressing the immune response in a subject requiring the suppression.  
 CC It is also useful for treating lymphoproliferative disorder, cancer e.g.  
 CC adenocarcinoma, leukaemia, myeloma, sarcoma, etc, neurodegenerative  
 CC disorder, inflammatory disorders e.g. rheumatoid arthritis, asthma,  
 CC multiple sclerosis, psoriasis, etc, neuronal disorders e.g. Alzheimer's  
 CC disease, dementia, depression, epilepsy, etc, immune disorder or immune  
 CC related disorders such as acquired immuno deficiency syndrome (AIDS),  
 CC allergy, anaemia, atopic dermatitis, diabetes mellitus, myocardial  
 CC infarction, etc, developmental disorders e.g. Cushing's syndrome, renal  
 CC tubular acidosis, gonadal dysgenesis, dysplasia, cataract, etc. The  
 CC present sequence is human RAT1d6 protein.  
 XX  
 SQ Sequence 422 AA;  
 Query Match 100.0%; Score 2229; DB 23; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-181;  
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MQQPQOQOQPGGQOLGGGGAAPGAGGGGPGGPGCLRRRLKLLSFHGRHFR 60  
 DB 1 MQQPQOQOQPGGQOLGGGGAAPGAGGGGPGGPGCLRRRLKLLSFHGRHFR 60  
 QY 61 IASACDELSCFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPI 120  
 DB 61 IASACDELSCFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPI 120  
 QY 121 VESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAEQ 180  
 DB 121 VESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAEQ 180  
 QY 181 TQEDVSDEDEEMPEDEDLDHVMKEEPAGCKKSEDDGIGKENLAILEKIKKNORQD 240  
 DB 181 TQEDVSDEDEEMPEDEDLDHVMKEEPAGCKKSEDDGIGKENLAILEKIKKNORQD 240  
 QY 241 YLNGAVSGSVQATDRLMKELRDYRSQSFSGNAYAVELVNDSDLYDNVVKLLKVDQDSALH 300  
 DB 241 YLNGAVSGSVQATDRLMKELRDYRSQSFSGNAYAVELVNDSDLYDNVVKLLKVDQDSALH 300  
 QY 301 NDLQILKEGADPILLNFSKDNFPDPPVPRVVSPLSGGYVLGGGATCMELLTKQGW 360  
 DB 301 NDLQILKEGADPILLNFSKDNFPDPPVPRVVSPLSGGYVLGGGATCMELLTKQGW 360  
 QY 361 SSAYSIESVIMQISATLVKGRARVQFGANKSQYSILTRAQOQSKYSLVQIHEKNGWYTPPKE 420  
 DB 361 SSAYSIESVIMQISATLVKGRARVQFGANKSQYSILTRAQOQSKYSLVQIHEKNGWYTPPKE 420

QY 421 DG 422  
 DB 421 DG 422  
 RESULT 2  
 AAG73684  
 ID AAG73684 standard; Protein; 374 AA.  
 XX  
 AC AAG73684;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Human colon cancer antigen protein SEQ ID NO:4448.  
 XX  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200122920-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000WO-US26524.  
 XX  
 PR 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 PI WPI; 2001-235357/24.  
 XX N-PSDB; AAH33115.  
 DR  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX  
 PS Claim 11; Page 6262-6264; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Pa,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAH7789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 SQ Sequence 374 AA;  
 Query Match 83.6%; Score 1863; DB 22; Length 374;  
 Best Local Similarity 98.4%; Pred. No. 8.6e-150;  
 Matches 358; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 59 FRIASACDELSCFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPI 118  
 DB 11 FRIASACDELSCFLLAGAGGAGAGAPCPASPTGSPGDPVRIHCNITESYPAPVPI 70  
 QY 119 WSVESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAE 178

Db 71 WSVSDDDNLAALVRLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAE 130  
QY 179 OCTQEDVSSDEDEMPEDTDLHYEMKEEPAEGKSEDDGIGKENLAILEKIKKNOR 238  
Db 131 OCTQEDVSSDEDEMPEDTDLHYEMKEEPAEGKSEDDGIGKENLAILEKIKKNOR 190  
QY 239 QDYLANGVSGVQATDRLMKELRDYRSQSFKGNYAVELVNDSDLYDNVKKLVQDQSA 298  
Db 191 QDYLANGVSGVQATDRLMKELRDYRSQSFKGNYAVELVNDSDLYDNVKKLVQDQSA 250  
QY 299 LNDLQILKEGADFIILNFSKDNFPDPFVRVSPVLSGGYVGGGGAICMELLTKQ 358  
Db 251 LNDLQILKEGADFIILNFSKDNFPDPFVRVSPVLSGGYVGGGGAICMELLTKQ 310  
QY 359 GWSAYSIESYIMQISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPP 418  
Db 311 GWSAYSIESYIMQISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPP 370  
QY 419 KEDG 422  
Db 371 KEDG 374

RESULT 3  
AA48312  
ID AA48312 standard; Protein; 291 AA.

XX AA48312;  
XX AC  
XX DT 08-DEC-1999 (first entry)  
XX DE Human prostate cancer-associated protein 9.

XX KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
XX KW cancer; tissue specificity; human.  
XX OS Homo sapiens.

XX DE19811194-A1.  
XX PD 16-SEP-1999.

XX PF 10-MAR-1998; 98DE-1011194.  
XX PR 10-MAR-1998; 98DE-1011194.

XX PA (META-) METAGEN GES. GENOMFORSCHUNG MBH.  
XX PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-519629/44.  
XX DR N-PSDB; AAZ33482.

XX PT New nucleic acid expressed at high level in normal prostatic tissue and  
XX PT encoded polypeptides, used to treat cancer and screen for therapeutic  
XX PT agents

XX PS Claim 22; 128; 194pp; German.

XX CC This invention describes novel nucleic acid sequences (A) that are  
XX CC expressed at high level in normal prostatic tissue. Polypeptides (I)  
XX CC encoded by (A) are used: (a) for identifying agents for treatment of  
XX CC prostatic cancer and (b) for therapy of prostate cancer, optionally  
XX CC where expressed by gene therapy methods. (A) is also used to isolate  
XX CC full-length genes (for gene therapy) and for recombinant production of  
XX CC (I), which can be used to raise specific antibodies. (A) are identified  
XX CC by assembly of ESTs (expressed sequence tags) before these are analyzed  
XX CC for expression pattern (tissue specificity). This approach eliminates  
XX CC many of the false results, as regards tissue specificity, associated  
XX CC with known methods that use single (usually short) ESTs. AA48304-Y48456  
XX CC represent peptides encoded by the expressed sequence tags described in  
XX CC the method of the invention.

SQ Sequence 291 AA;

Query Match 67.7%; Score 1509; DB 20; Length 291;  
Best Local Similarity 100.0%; Pred. No. 7.6e-120;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 LERLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDVSSDED 191  
Db 1 LERLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDVSSDED 60

QY 192 EEMPEDETDLDHYEMKEEPAEGKSEDDGIGKENLAILEKIKKNQDYLANGVSGSVQ 251  
Db 61 EEMPEDETDLDHYEMKEEPAEGKSEDDGIGKENLAILEKIKKNQDYLANGVSGSVQ 120

QY 252 ATDRLMKELRDYRSQSFKGNYAVELVNDSDLYDNVKKLVQDQSAHNDLQILKEG 311  
Db 121 ATDRLMKELRDYRSQSFKGNYAVELVNDSDLYDNVKKLVQDQSAHNDLQILKEG 180

QY 312 ADFILLNFSKDNFPDPFVRVSPVLSGGYVGGGGAICMELLTKQWSAYSIESYIM 371  
Db 181 ADFILLNFSKDNFPDPFVRVSPVLSGGYVGGGGAICMELLTKQWSAYSIESYIM 240

QY 372 QISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPPKEDG 422  
Db 241 QISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPPKEDG 291

RESULT 4  
ABB97360  
ID ABB97360 standard; Protein; 340 AA.

XX ABB97360;  
XX AC  
XX DT 27-JUN-2002 (first entry)  
XX DE Novel human protein SEQ ID NO: 628.

XX KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;  
XX KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
XX KW neuroprotective; antiparkinsonian; protein therapy; EST;  
XX KW expressed sequence tag.

XX OS Homo sapiens.

XX PN WO200222660-A2.

XX PD 21-MAR-2002.

XX PF 10-SEP-2001; 2001WO-US26015.

XX PR 11-SEP-2000; 2000US-0659671.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX DR WPI; 2002-292408/33.

XX DR N-PSDB; ABN32546.

XX PT An isolated polynucleotide for treating diseases associated with its  
XX PT encoded polypeptide such as cancer and multiple sclerosis -

XX PS Example 2; SEQ ID NO 628; 509pp; English.

XX CC The present invention provides the protein and coding sequences of 414  
XX CC novel human proteins. These were isolated from expressed sequences tags  
XX CC (ESTs). They can be used to stimulate cell growth, to regulate  
XX CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
XX CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
XX CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
XX CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
XX CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention.  
 XX  
 SQ Sequence 340 AA;

Query Match 58.38; Score 1300.5; DB 23; Length 340;  
 Best Local Similarity 67.88; Pred. No. 5.1e-102;  
 Matches 261; Conservative 30; Mismatches 41; Indels 53; Gaps 5;

QY 41 LRREKLLEIFRHRHERFRIASACLDSEFLLAGAGAGAGAAPGPH-LPPRGVPG 99  
 DQ 6 LKAEKFLASIFDKNHERFRVSKWLDLHCQFLVPOQG-----SPSLPP----- 51  
 QY 100 DPVRIHCNITESYPAVPIWNSVSDPNLAALVRLVDIKKNTLLQHLKRIISDLCKL 159  
 DQ 52 -PLTLHCNITESYPSPPFIWVDSNPATSVLERLEDTKNNL----- 94  
 QY 160 YNLPQHPDVMLOQPAEQCTQEDVSS--EDEDEEMPEDETLDDHYEMKEEPSAEGKKS 217  
 DQ 95 -----NCTTEVTSEEEEEEEEAEDIEDLHYEMKEEPSISGKKS 135  
 QY 218 EDDGIGKENLAILEKIKKNQODYLVNGAVSGSVQATDRLMKELDIYRSQSGGNVAVE 277  
 DQ 136 EDEGIEKENLAILEKIKKTQODHLNGAVSGSVQASDRLMKELDIYRSQSYKTGYSVE 195  
 QY 278 LVNSLDVNWVKKLVQDQDSALHNDLQILKEKEGADPILLNFSPKDFPDPPFVRVVP 337  
 DQ 196 LINDSLDWHVKKLVQDQDPSPLHSDLQILKEKEGIEYILLNFSPKDFPDPPFVRVLP 255  
 QY 338 VLSGGYVLGGAIACMELTQGWSSAYSIESVIMQISATLVKGRVQFGANKSOYSILTR 397  
 DQ 256 VLSGGYVLGGALCMELTQGWSSAYSIESVIMQINATLVKGRVQFGANKNQYNLAR 315  
 QY 398 AQQSKYSLVQTHKNGWYTPPKEDG 422  
 DQ 316 AQQSYNSIVQHKNGWYTPPKEDG 340

RESULT 5  
 AAE24494  
 ID AAE24494 standard; Protein; 245 AA.  
 XX  
 AC AAE24494;  
 XX  
 DT 04-OCT-2002 (first entry)  
 XX  
 DE Human RATL1d6 transmembrane domain.  
 XX  
 KW Human; ubiquitin conjugating enzyme; UBC; RATL1d6; immune disorder;  
 KW regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;  
 KW tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;  
 KW sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma;  
 KW multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia;  
 KW depression; epilepsy; acquired immuno deficiency syndrome; allergy;  
 KW AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;  
 KW myocardial infarction; renal tubular acidosis; gonadal dysgenesis;  
 KW dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV;  
 KW anticonvulsant; antiinflammatory; Cushing's syndrome; cardiatic;  
 KW ophthalmological; transmembrane domain.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200236741-A2.  
 XX  
 PD 10-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-US46559.  
 XX  
 PR 30-OCT-2000; 2000US-244688P.  
 PR 30-JUL-2001; 2001US-308706P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Bowen MA, Wu Y, Yang W, Finger JN;  
 XX WPI; 2002-479758/51.  
 XX  
 PT Novel ubiquitin conjugating enzyme polypeptide isolated from activated  
 PT human T cell, for screening modulators useful for treating cancer,  
 PT immune disorder, lymphoproliferative disorder, neurodegenerative  
 PT disorder  
 XX  
 PS Claim 6; Page 61; 169pp; English.  
 XX  
 CC The invention relates to a novel ubiquitin conjugating enzyme (UBC)  
 CC homologue, RATL1d6 (regulated in activated T-lymphocytes 1d6) and its  
 CC corresponding nucleic acid. The invention also relates to methods for  
 CC treating, diagnosing, preventing and screening for disorders related  
 CC to the expression of RATL1d6. UBC is useful for screening for candidate  
 CC compounds capable of binding to and/or modulating its activity. UBC is  
 CC useful for treating an immune or neuronal disorder in a mammal. The  
 CC method is useful for treating a cancer or tumour. It is useful for  
 CC suppressing the immune response in a subject requiring the suppression.  
 CC It is also useful for treating lymphoproliferative disorder, cancer e.g.  
 CC adenocarcinoma, leukaemia, myeloma, sarcoma, etc, neurodegenerative  
 CC disorder, inflammatory disorders e.g. rheumatoid arthritis, asthma,  
 CC multiple sclerosis, psoriasis, etc, neuronal disorders e.g. Alzheimer's  
 CC disease, dementia, depression, epilepsy, etc, immune disorder or immune  
 CC related disorders such as acquired immuno deficiency syndrome (AIDS),  
 CC allergy, anaemia, atopic dermatitis, diabetes mellitus, myocardial  
 CC infarction, etc, developmental disorders e.g. Cushing's syndrome, renal  
 CC tubular acidosis, gonadal dysgenesis, dysplasia, cataract, etc. The  
 CC present sequence is human RATL1d6 transmembrane domain.  
 XX  
 SQ Sequence 245 AA;

Query Match 58.0%; Score 1293; DB 23; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-101;  
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 PHLPGRGVPDPRVHCNITESYPAVPIWNSVSDPNLAALVRLVDIKKNTLLQH 148  
 DQ 1 PHLPGRGVPDPRVHCNITESYPAVPIWNSVSDPNLAALVRLVDIKKNTLLQH 60  
 QY 149 LKRIISDLCKLYNLPQHPDVMLOQPAEQCTQEDVSSDEDEEMPEDETLDDHYEMKE 208  
 DQ 61 LKRIISDLCKLYNLPQHPDVMLOQPAEQCTQEDVSSDEDEEMPEDETLDDHYEMKE 120  
 QY 209 EEPAGCKKSEDDGIGKENLAILEKIKKNQODYLVNGAVSGSVQATDRLMKELDIYRSQS 268  
 DQ 121 EEPAGCKKSEDDGIGKENLAILEKIKKNQODYLVNGAVSGSVQATDRLMKELDIYRSQS 180  
 QY 269 FKGGNTAVELVNSLDVNWVKKLVQDQDSALHNDLQILKEKEGADPILLNFSPKDFPD 328  
 DQ 181 FKGGNTAVELVNSLDVNWVKKLVQDQDSALHNDLQILKEKEGADPILLNFSPKDFPD 240  
 QY 329 PPFVR 333  
 DQ 241 PPFVR 245

RESULT 6  
 AAB41773  
 ID AAB41773 standard; Protein; 263 AA.  
 XX  
 AC AAB41773;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF1537 polypeptide sequence SEQ ID NO:3074.  
 XX  
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiatic;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;



antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX Homo sapiens.  
 OS  
 XX WO200058473-A2.  
 PN  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shinketsu RA, Leach M;  
 PI  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC75982.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease.  
 PS  
 XX Claim 11; Page 2294; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 XX Sequence 263 AA:  
 Query Match 56.8%; Score 1266; DB 21; Length 263;  
 Best Local Similarity 94.3%; Pred. No. 3e-99;  
 Matches 246; Conservative 4; Mismatches 9; Indels 2; Gaps 2;  
 QY 39 PCLRELKLLSIFRGRHERPRIASACLDLSCEFLLAGAGAGAGAAGPPLPRGSGVP 98  
 Db 2 PCLRELKLLSIFRGRHERPRIASACLDLSCEFLLAGAGAGAGAAGPPLPRGSGVP 61  
 QY 99 GDPVRIHCNITESYPVPPINSVESDDPNLAALVRLVDIKGNFTLLQHLKRIISDLCK 158  
 Db 62 GDPVRIHCNITESYPVPPINSVESDDPNLAALVRLVDIKGNFTLLQHLKRIISDLCK 121  
 QY 159 LYNLPQHPDVEMLDQPLPAEQCTQEDVSSEDEDEMPEDTDLHYENKEEPAEGKKSE 218

Db 122 LYNLPQHPDVEMLDQPLPAEQCTQEDVSSEDEDEMPEDTDLHYENKEEPAEGKKSE 181  
 QY 219 DDGIGKENLAILEIKKRNQDYLVNGAVSGVQATDRLMKELRD-IVRSQSFKGNY-AV 276  
 Db 182 DDGIGKENLAILEIKKRNQDYLVNGAVSGVQATDRLMKELQGVITXSQSFKGNYXSS 241  
 QY 277 ELVNDSLYDNVNVKLLKVDQDS 297  
 Db 242 NSWNSLYGWDVQLLKVDQGS 262  
 RESULT 7  
 AAY73387  
 ID AAY73387 standard; Protein; 308 AA.  
 XX  
 AC AAY73387;  
 XX  
 DT 24-FEB-2000 (first entry)  
 XX  
 DE HTRM clone 3340290 protein sequence.  
 XX  
 KW HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
 KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;  
 KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
 KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9957144-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 04-MAY-1999; 99WO-US09935.  
 XX  
 PR 05-MAY-1998; 98US-0084254.  
 PR 07-AUG-1998; 98US-0095827.  
 PR 02-OCT-1998; 98US-0102745.  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;  
 PI Gerstin EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;  
 DR  
 DR WPI; 2000-052941/04.  
 N-PSDB; AAZ52472.  
 XX  
 PT New peptides useful for diagnosis, prevention and treatment of cancer  
 PT and immune disorders.  
 XX  
 PS Claim 1; Page 144-145; 193pp; English.  
 XX  
 CC AAY73325-V73389 are human transcriptional regulator molecule (HTRM)  
 CC protein sequences. The HTRM protein and nucleotide sequences are useful  
 CC for preventing or treating disorders associated with decreased expression  
 CC or activity of HTRM which include cell proliferative disorders such as  
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
 CC leukaemia; immune disorders such as AIDS, Addison's disease, diabetes  
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
 CC of the HTRM polypeptides are useful for treating or preventing disorders  
 CC associated with increased expression or activity of HTRMs. HTRM  
 CC polypeptides, their immunogenic fragments or oligopeptides are useful  
 CC for screening libraries of compounds in drug screening techniques.  
 CC Polynucleotides encoding HTRM are useful for blocking the transcription  
 CC of mRNA and regulating gene function by modulating the activity of  
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or  
 CC treat disorder associated with decreased HTRM expression. Antibodies  
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
 CC for diagnosing disorders associated with the expression of HTRM.  
 CC particularly in assays that detect the expression of HTRM. Nucleotide  
 CC sequences encoding HTRM may be useful to generate hybridization probes  
 CC useful in mapping the naturally occurring genomic sequence and to detect

CC differences in gene sequences among normal, carrier and affected  
 CC individuals. Using diagnostic assays, cancer can be detected prior to  
 CC the appearance of clinical symptoms and thereby progression of cancer can  
 CC be prevented by aggressive treatment or preventive measures.

XX  
 SQ Sequence 308 AA;  
 Query Match 50.3%; Score 1120.5; DB 21; Length 308;  
 Best Local Similarity 55.7%; Pred. No. 9e-87;  
 Matches 230; Conservative 29; Mismatches 41; Indels 85; Gaps 6;  
 QY 41 LRRLKLESTFHRGHERFRFASACLDCEFLLAGAGAGAAAPGPHLPGRGSPVG 99  
 DB 6 LKAEKFLASFDKXNHERFRVSWKLDLHCQFLVPOQG-----SPHSLPP----- 51  
 QY 100 DPVRIHCNITESYPAPPIWVSDDPNLAALVRLVDIKKGNFLLQLHLKRIISDLCKL 159  
 DB 52 -PLTLHCNITESYPSPINFDSEDPNLASVLERLEDTKNNL----- 94  
 QY 160 YNLPHQPDVEMLOPLPAEQCTQEDVSS--EDEDEMPEDTEDLDHYEMKEEPEAGKKS 217  
 DB 95 -----NCTTEVTSEEEEEEAEADIEDLDHYEMKEEPEISGKKS 135  
 QY 218 EDDGIGKENLAILEKIKKQKQDYLNGAVSGSVQATDRLMKELRDIYRSQFKGNYAVE 277  
 DB 136 EDEGIGKENLAILEKIKKQKQDYLNGAVSGSVQATDRLMKELRDIYRSQFKGNYAVE 195  
 QY 278 LVNDSLDVNWVKLKQVQDSALHNDLQILKEKEGADFIILNFSKDNFPDPPFVRVVP 337  
 DB 196 LINDSLVDHWVKQKQVDPSPHSDQLQILKEKEGIEVILLNFSKDNFPDPPFVRVVP 255  
 QY 338 VLSGGYVLGGGAIQCMELLTKQGWSSAYSIESVIMQISATLVKGRVQFGANKSQYSLTR 397  
 DB 256 VLSGGYVLGGGAIQCMELLTKQ-----NQYNLAR 283  
 QY 398 AQQYSKSIQVQIHEKNGWTPPKEDG 422  
 DB 284 AQQYSKSIQVQIHEKNGWTPPKEDG 308

## RESULT 8

ABB59516  
 ID ABB59516 standard; Protein; 397 AA.  
 AC ABB59516;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5340.  
 XX Drosophila melanogaster polypeptide; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 FI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL03619.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

## Interactions -

PT  
 XX Disclosure; SEQ ID NO 5340; 21pp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 397 AA;

Query Match 45.8%; Score 1020.5; DB 22; Length 397;  
 Best Local Similarity 49.0%; Pred. No. 4.1e-78;  
 Matches 202; Conservative 72; Mismatches 87; Indels 51; Gaps 4;

QY 41 LRRELKLESIPHRGHERFRFASACLDCEFLLAGAGAGAGAAAPGPHLPGRGSPVG 100  
 DB 7 LKQEIITLEKIFPKNHERFOILNSSVDELLCRI-----DKNGK 45  
 QY 101 PVRHCNITESYPAPPIWVSDDPNLAALVRLVDIKKGNFLLQLHLKRIISDLCKLY 160  
 DB 46 RYDIHANITETYPSSPPVWFAESEEVTNAVQILSNTNGRDNHVINQVIGILLRELCLRH 105  
 QY 161 NLPQHPDVEMLDPLPAEQCTQEDVSEDE-----DEEMPEDEDLDHYEM 206  
 DB 106 NVLPDIDNIALPLQTPPPSPRCEQRPGGGAGGGGPHGNEETSDQEEIDPIG 165  
 QY 207 KEEPEAGKK-----SEDDGIGKENLAILEKIKKQKQDYLNGAVSGSVQATD 254  
 DB 166 ESEQSEGEDDLPLEMDVYRSTSKDDMEVEHLATLEKLRQSDYLAGSVSGSVQATD 225  
 QY 255 RLMKELRDIYRSQFKGNYAVELVNDSLVDWVKKLVQDQDSALHNDLQILKEKEGADF 314  
 DB 226 RLMKELRDIYRSQFKGNYAVELVNDSLVDWVKKLVQDQDSALHNDLQILKEKEGADF 285  
 QY 315 ILLNFSKDNFPDPPFVRVVPVLSGGYVLGGGAIQCMELLTKQGWSSAYSIESVIMQIS 374  
 DB 286 ILLNLFKETYPFPPFVRVVPVHISGGYVLGGGAIQCMELLTKQGWSSAYSIESVIMQIS 345  
 QY 375 ATLVKGRVQFGANKS-----QYSLTRAQGSYKSLVQIHEKNGWTPPKEDG 422  
 DB 346 ATLVKGRVQFGANKS-----QYSLTRAQGSYKSLVQIHEKNGWTPPKEDG 397

## RESULT 9

AAE24524  
 ID AAE24524 standard; Protein; 164 AA.  
 XX  
 AC AAE24524;  
 XX  
 DT 04-OCT-2002 (first entry)  
 DE Human RAT1d6 ubiquitin conjugating enzyme (UBC) domain.  
 XX  
 KW Human; ubiquitin conjugating enzyme; UBC; RAT1d6; immune disorder;  
 KW regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;  
 KW tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;  
 KW sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma;  
 KW multiple sclerosis; psoriasis; neuroendocrine; Alzheimer's disease; dementia;  
 KW depression; epilepsy; acquired immunodeficiency syndrome; allergy;  
 KW AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;  
 KW myocardial infarction; renal tubular acidosis; gonadal dysgenesis;  
 KW dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV;  
 KW anticonvulsant; antiinflammatory; Cushing's syndrome; cardiac;  
 KW ophthalmological; transmembrane domain.







Db 124 ESYSSSPWFVDSDELNTSVLECLD-DKNNLLHQQLKWLICGLCRLYNLPKHLDA 182



## RESULT 2

A41241  
ubiquitin-conjugating enzyme CDC34 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein D4211; protein YRD295; protein YD9609.08c; protein YDR054c;  
C:Species: Saccharomyces cerevisiae  
C>Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 21-Jul-2000  
C:Accession: A41241; S54038; S58834; S61743; S67870  
R:Goebel, M.G.; Yochem, J.; Jentsch, S.; McGrath, J.P.; Varshavsky, A.; Byers, B.  
Science 241, 1331-1335, 1988  
A:Title: The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme.  
A:Reference number: A41241; MUID:88321694; PMID:2842867  
A:Accession: A41241  
A:Molecule type: DNA  
A:Residues: 1-295 <GOE>  
A:Cross-references: GB:M21877; NID:gl73123; PIDN:AAA35188.1; PID:gl73124  
A>Note: the authors translated the codon ATA for residue 50 as Ala  
R:Hunt, S.; Bowman, S.; Harris, D.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54038  
A:Accession: S54038  
A:Molecule type: DNA  
A:Residues: 1-295 <HUN>  
A:Cross-references: EMBL:Z49209; NID:g798897; PID:g798905  
R:Brandt, P.; Otto, B.; Ramlow, S.; Blocker, H.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S58832  
A:Accession: S58832  
A:Molecule type: DNA  
A:Residues: 1-295 <BRA>  
A:Cross-references: EMBL:X84162; NID:g706817; PID:g706820  
R:Brandt, P.; Ramlow, S.; Otto, B.; Blocker, H.  
Yeast 12, 85-90, 1996  
A:Title: Nucleotide sequence analysis of a 32,500 bp region of the right arm of Sacchar  
A:Reference number: S61741; MUID:96381250; PMID:8789263  
A:Accession: S61743  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-295 <BRW>  
A:Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58970.1; PID:g706820  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995  
R:Blocker, H.; Brandt, P.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67587  
A:Accession: S67870  
A:Molecule type: DNA  
A:Residues: 1-295 <BLO>  
A:Cross-references: EMBL:Z74350; NID:gl431496; PID:e253311; PID:gl431497; MIPS:YDR054c  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:CDC34; UBC3  
A:Cross-references: SGD:S0002461; MIPS:YDR054c  
A:Map position: 4R  
C:Keywords: nucleus

Query Match 6.1%; Score 135.5; DB 2; Length 295;  
Best Local Similarity 21.4%; Pred. No. 0.19;  
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;  
QY 273 NYAVELVNDLS-LYDWNVLLKVDQDSALHNDLIQLEKAGDFILLNFSKDNFPFPPFP 331  
DB 28 SFHIELEDDSNFTWNGVMVLNEDSIYH-----GGFFKAQMRFPEDFPFSPQ 76  
QY 332 VRVSPVLGGVGLGGGAICMELTFKOG-----WSSAYSIESVIMQISATL--- 377  
DB 77 FR-FTPAYHNPNVYDRGLICISILHQSGDPMTDEPDAETSPVQVESVLISVLSLEBDP 135  
QY 378 -----VRKARYQFGANKSQYS-----LTRAQQYSKSLVQTHE--- 410  
DB 136 NINSPANVDAADYRKNEPQYKQVKMEVRSKQDIPKGFIMPTSESAYISQSKLDEPES 195  
QY 411 -----KNGWYTPPKED 421

Db 196 NKDMADNFWYDSLDLD 211

## RESULT 3

T03166  
probable immediate early protein - alcelaphine herpesvirus 1  
C:Species: alcelaphine herpesvirus 1  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Nov-1999  
C:Accession: T03166  
R:Essner, A.; Pflanz, R.; Fleckenstein, B.  
J. Virol. 71, 6517-6525, 1997  
A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.  
A:Reference number: Z14840; MUID:97404639; PMID:9261371  
A:Accession: T03166  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1300 <ENS>  
A:Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAC58118.1; PID:g2338034  
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo  
Query Match 6.0%; Score 133; DB 2; Length 1300;  
Best Local Similarity 24.4%; Pred. No. 2;  
Matches 65; Conservative 32; Mismatches 103; Indels 66; Gaps 10;  
QY 6 PQGQQQPGP-----GQQL-----GGGAAPGAGGGPGGP-GPGPCLRLRELKLESIFHRG 55  
DB 646 PEGECQSPSCGQVPGDPGPEGSSGSGSGEGSPSGS-----SEG 692  
QY 56 HERFRASACLDLSCFLLAGAGAGAGAAAP-----GPHLPGRGVPDVR 103  
DB 693 QQVPKGAEGSEGGPCR-----PGPDDEGDPGPGTEGEGCGGPDGDPGPGP--- 745  
QY 104 IHCNIETSPAVPIINVSDDPNLAALVRLVDIKKNTLLIQLHLK----- 150  
DB 746 ---DGTGEGPCGPGGDEGDPESEGTED--DIKVLTELIGSMKLDSSDSDSSDSSD 800  
QY 151 ----RIISDLKLYNLPHDPVEMLQPLPAEQCTQE--DVSEDEDEEMPEDELDHY 204  
DB 801 SANRRALGVCGSHSSKSDDEEEEEEEDDEDEDEDEDEDEDEDEDEDEDEDEDE 860  
QY 205 EMKEEPAECKKSEDDGIGKENLAIL 230  
DB 861 DEEEEDDEEEEDDEEEEEEVII 886

## RESULT 4

S39580  
HERM protein - human  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 02-Aug-2002  
C:Accession: S39580  
R:Muchardt, C.; Yaniv, M.  
EMBO J. 12, 4279-4290, 1993  
A:Title: A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm 9  
A:Reference number: S39580; MUID:94038910; PMID:8223438  
A:Accession: S39580  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1586 <MUC>  
A:Cross-references: EMBL:X72889; NID:g414116; PIDN:CAA51407.1; PID:g414117  
C:Superfamily: human SNF2alpha protein; bromodomain homology  
F:216-249/Region: glutamine-rich  
F:1423-1478/Domain: bromodomain homology <BRO>  
Query Match 5.9%; Score 132.5; DB 2; Length 1586;  
Best Local Similarity 22.0%; Pred. No. 2.8;  
Matches 68; Conservative 39; Mismatches 109; Indels 93; Gaps 14;  
QY 3 QPQPQGOQP-----GPGQOLGGOGA-----APGAGGGGPGGPGPCLRLRELKLL 48  
DB 241 QPOTQOQQQALVNNRNPSPGPGSELGSPSTPQKLVP-----AFGRRSPAP----- 287



QY 49 ESIFRHERFRIASACLDLSCFLLAGAGAGAGAAPGPHLPGRGVPDPVRIHCNI 108  
 Db 288 -----PAAQPPAAAVPGSVPO--PAPQPSVP-LQL 317  
 QY 109 TESYPVPPPIWSVESDDNLAALVLRVLDIKKNTLLQHLKRIISDLCKLYNLPHQ--P 166  
 Db 318 QOKOSRISPIOKPQGLDP-----VEILQREYRLQ--ARIAHRIQELNLPGLSLPP 366  
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSEDEEMPEDETLQHYEMKEEPAEGKKE 218  
 Db 367 DLRTKATVELKALRLNLFQRLQREEVVACMRDITLETALNSKAYKRSKQTLREARMTE 426  
 QY 219 DDGIGKENLAILEKIKKNQDYLNAGVSGSVQATDRMLKELRDYRSQSFKGGNYAVEL 278  
 Db 427 K--LEKQOKIEOERKRQKHQEYLSILOHA-----KDFKEYHRSVAGK-----IQK 471  
 QY 279 VNDSLYDWN 287  
 Db 472 LSKAVATWH 480

## RESULT 5

T08465  
 ubiquitin-protein ligase (EC 6.3.2.19) - fruit fly (*Drosophila melanogaster*)  
 N:Alternate names: ubiquitin conjugating enzyme  
 C:Species: *Drosophila melanogaster*  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 03-Jun-2002  
 C:Accession: T08465  
 R:Kirby, R.J.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z16421  
 A:Accession: T08465  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-199 <KIR>

A:Cross-references: EMBL:X92838; NID:g1359613; PID:g1054722

A:Experimental source: strain Canton-S

C:Genetics:

A:Gene: UbcD4

A:Cross-references: FlyBase:FBgn0015321

C:Keywords: yeast ubiquitin-protein ligase UBC1

C:Keywords: cell cycle control; DNA repair; ligase; protein degradation

Query Match 5.9%; Score 131; DB 2; Length 199;  
 Best Local Similarity 24.7%; Pred. No. 0.21;  
 Matches 38; Conservative 32; Mismatches 56; Indels 28; Gaps 7;

QY 252 ATDRMLKELRDYRSQSFKGGNYAVELVNDSLYDWNVKKLLKVDODSALHNDLQILKEK-- 309  
 Db 5 AVSRIKREFEVMRSEIEVQCSIKIELVNDSD--W-----TELGEIAGPPDTPY 51  
 QY 310 EGADFILLNFSKDNFPDPFVRVGVPLSGGVVLGGGATCMELLTKQGWSSAYSIESV 369  
 Db 52 EGGKFV-LEIKVPETYPFNPFPKARFIRIWHNPNISSVTGAICLDIL-KDNWAAAMTLRTV 109  
 QY 370 IMQISATLVKGG-----ARVQFGANKSOYSL 395  
 Db 110 LLSLQALLAAEPPDPQDAVVAYQF---KDKIDL 140

## RESULT 6

SNF2alpha protein - human  
 C:Species: *Homo sapiens* (man)  
 C>Date: 10-Dec-1994 #sequence\_revision 10-Nov-1995 #text\_change 02-Aug-2002  
 C:Accession: S45251  
 R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.  
 Nucleic Acids Res. 22, 1815-1820, 1994  
 A:Title: Two human homologues of *Saccharomyces cerevisiae* SWI2/SNF2 and *Drosophila* brahm  
 A:Reference number: S45251; MUID:94268902; PMID:8208605  
 A:Accession: S45251  
 A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-1572 <CHI>  
 A:Cross-references: GB:D26155; NID:g505086; PIDN:BAA05142.1; PID:d1005684; PID:g9876  
 C:Superfamily: human SNF2alpha protein; bromodomain homology  
 F:1409-1464/Domain: bromodomain homology <BRO>

Query Match 5.8%; Score 129.5; DB 2; Length 1572;  
 Best Local Similarity 21.7%; Pred. No. 4.3;  
 Matches 67; Conservative 40; Mismatches 109; Indels 93; Gaps 14;  
 QY 3 QPQGGQGOQP-----GPGQOLGGOGA-----APGAGGGGGGPGPGPCLRLRELKLL 48  
 Db 245 QPOTQOQQPALVNYNRPSPGPGSPSTPQKLPVP---APGGRSPAP----- 291  
 QY 49 ESIFRHERFRIASACLDLSCFLLAGAGAGAGAAPGPHLPGRGVPDPVRIHCNI 108  
 Db 292 -----PAAQPPAAAVPGSVPO--PAPQPSVP-LQL 321  
 QY 109 TESYPVPPPIWSVESDDNLAALVLRVLDIKKNTLLQHLKRIISDLCKLYNLPHQ--P 166  
 Db 322 QOKOSRISPIOKPQGLDP-----VEILQREYRLQ--ARIAHRIQELNLPGLSLPP 370  
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSEDEEMPEDETLQHYEMKEEPAEGKKE 218  
 Db 371 DLRTKATVELKALRLNLFQRLQREEVVACMRDITLETALNSKAYKRSKQTLREARMTE 430  
 QY 219 DDGIGKENLAILEKIKKNQDYLNAGVSGSVQATDRMLKELRDYRSQSFKGGNYAVEL 278  
 Db 431 K--LEKQOKIEOERKRQKHQEYLSILOHA-----KDFKEYHRSVAGK-----IQK 475  
 QY 279 VNDSLYDWN 287  
 Db 476 LSKAVATWH 484

## RESULT 7

B70694  
 Probable InfB - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: *Mycobacterium tuberculosis*  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 02-Feb-2001  
 C:Accession: B70694  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete gene  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: B70694  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-900 <COL>  
 A:Cross-references: GB:281331; GB:ALI23456; NID:g3261650; PIDN:CAB03670.1; PID:g1648E  
 A:Experimental source: strain H37RV  
 C:Genetics:

A:Gene: infB  
 C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu  
 C:Keywords: GTP binding; nucleotide binding; P-loop  
 F:399-512/Domain: translation elongation factor Tu homology <ETU>  
 F:405-412/Region: nucleotide-binding motif A (P-loop)  
 F:509-512/Region: GTP-binding NKXD motif  
 F:545-547/Region: GTP-binding SAK/L motif  
 F:411,412,432,509,510,512,545/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser

Query Match 5.5%; Score 123.5; DB 2; Length 900;  
 Best Local Similarity 20.5%; Pred. No. 4.9;  
 Matches 92; Conservative 63; Mismatches 128; Indels 165; Gaps 21;

QY 4 PQQGGQ-QPGPQ-----OLGGQGAAPGAG--GGPGGGPGPGPCLRLRELKLLSIF 52  
 Db 211 PRPGGRGAPGAGCRSDAGGNGYRGVGAAPGTGFRGPGGGGGGRRGQ----- 260  
 QY 53 HRGHERFRIASACLDLSCFLLAGAGAGAGAAPGPHLPGRGVPDPVRIHCNI 95

R;Doughlas, S.; Zauner, S. 2001  
 Nature 410, 1091-1096, 2001  
 A:Title: The highly reduced nucleus of an enslaved algal nucleus.  
 A:Reference number: A99082; PMID:11323671  
 A:Accession: F90118  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-147 <DOU>  
 A:Cross-references: GB:AF083031; NID:g5690139; PIDN:AAF24004.1; GSPDB:GNO0152  
 C:Genetics:  
 A:Gene: ubc4  
 A:Map position: 3  
 A:Genome: nucleomorph  
 C:Superfamily: human ubiquitin-protein ligase E2  
 C:Keywords: nucleomorph  
  
 Query Match 5.5%; Score 121.5; DB 2; Length 147;  
 Best Local Similarity 25.3%; Pred. No. 0.56;  
 Matches 39; Conservative 29; Mismatches 59; Indels 27; Gaps 7



A:Experimental source: clone E02A10

C:Genetics:

A:Gene: CESP:E02A10.2

A:Map position: 5

A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 5.4%; Score 120.5; DB 2; Length 385;  
Best Local Similarity 22.9%; Pred. No. 2.4;  
Matches 54; Conservative 28; Mismatches 101; Indels 53; Gaps 6;  
QY 13 GPGQOLGGGGAAGAGGGGPGPCPLRELKLLSFHGRHFRFASACIDELSC 72  
Db 151 GGGCGGGGGGGGGGGGGGGGGGGG----- 178  
QY 73 FLGAGAGGAGAGPAGPHLPGRGVPDPVRIHGNITSPVAVPIWVESDDPNLA 132  
Db 179 ---GGGGGGGGGGGKREAINVVGHDDLKCNNEEL-----RIILNENTK 230  
QY 133 ERLVDIKKGNL-----OHLKRIISDLKLYNLPOHPDVEMLDQPLAEQ 186  
Db 231 KTLKEKVAQDYIVVCKNEKPAPFTAEITDDFCLOKENVHCTILRINHKE 289  
QY 187 SED--EDREMPEDTDLHYEMKEEPAEGKSEDDGIGKENLAILKIKKN 240  
Db 290 EEPKKEEKEEVEKDEKDEEP---KKEEKEEKEEVEKKEEKKD 341

#### RESULT 15

C86725

chromosome segregation SMC protein [imported] - Lactococcus lactis subsp. lactis (strain C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: C86725

R:Bolotin, A.; Wincker, P.; Mauder, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: C86725

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-924 <STO>

A:Cross-references: GB:AE005176; PID:gl2723724; PIDN:AAK04901.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: smc

Query Match 5.4%; Score 120.5; DB 2; Length 924;  
Best Local Similarity 24.2%; Pred. No. 7.8;  
Matches 60; Conservative 41; Mismatches 86; Indels 61; Gaps 13;

QY 127 NLAAYLE---RLVDIKKGNL---OHLKRIISDL-CKLYNLQHPDVEMLDQPLAE 178  
Db 329 NLKSRLEITQONLAEVQKSELLTEKNDLKLTLSEADLATLSEPEVVM--ERL 386  
QY 179 --OCTQEDVSEDEBEMPEDETDLDHYEMKEEPEAGKSEDDGIGKENLAIL 235  
Db 387 FVQLVEEAQISNEIVRNKAETILSR-----RQSEQDESRENLTKEKISOD 435  
QY 236 -NORODYLNGAVSGVQATDRLMKELDIYRSQFKGNYAVE-----LVNDSL 287  
Db 436 LSEAGENL-----TVKKEITLL--AKFEENQAEKKQAELELAQNKMYYD- 481  
QY 288 VKLLKVDQDSALHNDLQILKE-----GAQFILLNFS-----FKDNFPF 334  
Db 482 --LOELNQHKARLTSLQNIRESHNLFAQVAVMNASQIGVGIVGVSDDLTF 539  
QY 335 VSPVLGG 342  
Db 540 IDIALGG 547

Search completed: July 29, 2003, 13:47:59  
Job time : 44 secs



```

CC -----
DR EMBL; M21877; AAA35188.1; -
DR EMBL; X84162; CAA58970.1; -
DR EMBL; 274350; CAA98872.1; -
DR EMBL; 249209; CAA89083.1; -
DR PIR; A41241; A41241.
DR HSP; Q02159; 20C2.
DR SGD; S0002461; CDC34.
DR GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IPI.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
KW Ub1 conjugation pathway; Ligase; DNA replication; Nuclear protein;
KW Cell cycle; Cell division; Multigene family.
FT BINDING 95 95 UBIQUITIN (BY SIMILARITY).
FT DOMAIN 191 289 ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 295 AA; 34064 MW; 1CE3E0C3AB1436DC CRC64;
SQ
Query Match 6.1%; Score 135.5; DB 1; Length 295;
Best Local Similarity 21.4%; Pred. No. 0.11;
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;
QY 273 NYAVELVDS-LYDNNVKLLKVDQDSALHNDLQLKEGADFILLNFSFKDNFPDPFP 331
Db 28 SFHLEDDSNFTWNGVWLNEDSIYH-----GGFFKAQWRFEDFPSPQ 76
QY 332 VRVVSPVLGGVYLGGAICMELITKQG-----WSAYSIESVIMQISATL--- 377
Db 77 FR-FTPAIYHPNVDYRGLCSILHSGDPMTPDEPDAETSPVQVESLVISVLLSDP 135
QY 378 -----VKGARVQFCANKSQYS-----LTRAQSQYSKLVQIHE--- 410
Db 136 NINSPANTVDAVDYKKNPEYQKRVKMEVRSKQDIPKGFIMPTSESAYISOSKLDEPES 195
QY 411 -----KNGWYTPPKED 421
Db 196 NKDMADNEWYDSLDLD 211
RESULT 2
ID SN22_HUMAN STANDARD; PRT; 1586 AA.
AC P51531;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Possible global transcription activator SNF2L2 (SNF2-alpha).
GN SMARCA2 OR SNF2L2 OR BRM OR SNF2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94038910; PubMed=8223438;
RA Muchardt C., Fanly M.;
RT "A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and
RT Drosophila brm genes potentiates transcriptional activation by the
RT glucocorticoid receptor.";
RL EMBO J. 12:4279-4290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94268902; PubMed=8208605;
RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brhma are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor.";
RL Nucleic Acids Res. 22:1815-1820(1994).

```

```

CC -----
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId-P51531-1; Sequence=Displayed;
CC Name=Short;
CC IsoId-P51531-2; Sequence=VSP_000577;
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X72889; CAA51407.1; -
CC EMBL; D26155; BAA05142.1; -
CC PIR; S39580; S39580.
CC Genew; HGNC:11098; SMARCA2.
CC MIN; 600014; -
CC GO; GO:0005654; C:nucleoplasm; TAS.
CC GO; GO:0003713; F:transcription co-activator activity; TAS.
CC GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; TAS.
CC InterPro; IPR005576; BRK.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR006562; HSA.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00439; bromodomain; 1.
CC Pfam; PF00271; Helicase_C; 1.
CC Pfam; PF00176; SNF2_N; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00592; BRK; 1.
CC SMART; SM00297; BROMO; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICC; 1.
CC SMART; SM00573; HSA; 1.
CC PROSITE; PS00633; BROMODOMAIN_1; 1.
CC PROSITE; PS0014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT NP_BIND 745 752 ATP (POTENTIAL).
FT SITE 847 850 DEGH BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 BROMODOMAIN.
FT DOMAIN 1514 1525 POLY-GLU.
FT VARSPLIC 1397 1414 Missing (in isoform Short).
FT FTID=VSP_000577.
FT CONFLICT 239 239 P -> PQQP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 D -> H (IN REF. 2).
FT CONFLICT 1390 1390 C -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537F2A392807A CRC64;
Query Match 5.9%; Score 132.5; DB 1; Length 1586;
Best Local Similarity 22.0%; Pred. No. 1.6;
Matches 68; Conservative 39; Mismatches 109; Indels 93; Gaps 14;
QY 3 QPQPGQQPQ-----GPGQQLGCGA-----APGAGGGGCGGPGPCLRLKLL 48
Db 241 QPQTQQQQPALVNNRPSGPGPGLSPQKLPVP-----APGGRSPAP----- 287

```

QY 49 ESIFRHERPRIASACDELSCFELLAGAGGAGAGPGLPRLPRGSPVGDVRIHCNI 108  
 Db 288 -----PAAQPPAAAVPGSPVQ--PAPGQSPV-LQL 317  
 QY 109 TESTPVPPIWVESDDPNLAALVRLVDIKKGNLTLLOHLKRIISDLCKLYNLPQH--P 166  
 Db 318 QOKQSRISPIOKPQGLDP-----VEILQEREYRLQ--ARIAHRIQELNLPGLSPP 366  
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSDEDEMPEDFDLDHYEMKEEPAEGKKSE 218  
 Db 367 DLRTKATVELKALRLNLFQRLQROEVVACMRDPTLETALNSKAYKRSKQTLREARWTE 426  
 QY 219 DDGIGKENLAILEIKKKNORDYLNAGVSGSVQATRLMKELRLDIYSQSFKGNYAVEL 278  
 Db 427 K--LEKQOKEQEQRKQKQOYLNLSIQH-----KDFEYHRSVAGK-----IQK 471  
 QY 279 VNDSLYDWN 287  
 Db 472 LSKAVATWH 480

RESULT 3  
 UB6B\_MOUSE  
 ID UB6B\_MOUSE STANDARD; PRT; 207 AA.  
 AC P52483; O09180;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ubiquitin-conjugating enzyme E2-23 kDa (EC 6.3.2.19)  
 GN (Ubiquitin-protein ligase) (Ubiquitin carrier protein).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Swiss;  
 RX MEDLINE=96162026; PubMed=8576256;  
 RA Matuschewski K., Hauser H.P., Treier M., Jentsch S.;  
 RT "Identification of a novel family of ubiquitin-conjugating enzymes  
 RL with distinct amino-terminal extensions.";  
 RN J. Biol. Chem. 271:2789-2794 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Swiss;  
 RX MEDLINE=99087339; PubMed=9872334;  
 RA Pestov D.G., Grzeszkiewicz T.M., Lau L.F.;  
 RT "Isolation of growth suppressors from a cDNA expression library.";  
 RL Oncogene 17:3187-3197 (1998).  
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER  
 CC PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine - AMP +  
 CC diphosphate + protein N-ubiquityllysine.  
 CC -1- PATHWAY: Ubiquitin conjugation; second step.  
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
 CC thiolester formation (By similarity).  
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X92664; CAA63352.1; --  
 CC EMBL: AF003346; AAB60948.1; --  
 CC HSSP: P15731; IQCQ.  
 CC MGD: MGI:107412; Ubce4.  
 CC InterPro: IPR000608; UbQ\_conjugat.

DR Pfam: PF00179; UQ\_con; 1.  
 DR Prodom: PD000461; UbQ\_conjugat; 1.  
 DR SMART: SM00212; UBCC; 1.  
 DR PROSITE: PS00183; UBIQUITIN\_CONJUGAT\_1; 1.  
 DR PROSITE: PS0127; UBIQUITIN\_CONJUGAT\_2; 1.  
 KW Ubl conjugation pathway; Ligase; Multigene family.  
 FT BINDING 145 145 UBIQUITIN (BY SIMILARITY).  
 FT CONFLICT 31 31 K -> E (IN REF. 2).  
 SQ SEQUENCE 207 AA; 22911 MW; 86BCB5C8A72763F CRC64;

Query Match 5.8%; Score 129.5; DB 1; Length 207;  
 Best Local Similarity 24.2%; Pred. No. 0.17; 93; Indels 39; Gaps 10;  
 Matches 53; Conservative 34; Mismatches 93; Indels 39; Gaps 10;

QY 187 SEDEEMPEDETDLDHYEMKEEPAEGKKSEDDGIGKENLAILEIKKKNORDYLNAGV 246  
 Db 8 SDESPSTSGSSDAD-----QRDPAAPKPEEQ-----EERKPSATQOKKNTKL 51  
 QY 247 SGSVQA-----TDRLMKELRLDIYSQSFKGNYAVELVNDSLYDWNVLLKLVDDQSALHN 301  
 Db 52 SSKTTAKLSTSAKRIQKELAEITLDP-----PNCAGPKGDNIEYWKSTIL--GPPGSVY- 105  
 QY 302 DLQILKEGADFIILNFSFKDNFPDPFVRVSPVLSSGGVVLGGGATCMELLTKQGS 361  
 Db 106 -----EGGVF--FLDITFSSDYPPKPKVTFRTIYHCN--INSQGVICLDIL-KDNWS 154  
 QY 362 SAYSIESVIMQISATLVK-GKARVQFGANKSQYSITRAQ 399  
 Db 155 PALTISKVLLSCSLTDCNPADPLVGSIAQTQVLTNRAE 193

RESULT 4  
 SA3\_RAT  
 ID SA3\_RAT STANDARD; PRT; 1256 AA.  
 AC Q99M76;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cohesin subunit SA-3 (Stromal antigen 3) (SCC3 homolog 3).  
 GN STAG3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar;  
 RX MEDLINE=21482572; PubMed=11599053;  
 RA Bayes M., Prieto I., Noguchi J., Barbero J.L., Perez Jurado L.A.;  
 RT "Evaluation of the Stag3 gene and the synaptonemal complex in a rat  
 RL model.(as/as) for male infertility.";  
 RL Mol. Reprod. Dev. 60:414-417 (2001).  
 CC -1- FUNCTION: Meiosis specific component of cohesin complex. The  
 CC cohesin complex is required for the cohesion of sister chromatids  
 CC after DNA replication. The cohesin complex apparently forms a  
 CC large proteinaceous ring within which sister chromatids can be  
 CC trapped. At anaphase, the complex is cleaved and dissociates from  
 CC chromatin, allowing sister chromatids to segregate. The meiosis-  
 CC specific cohesin complex probably replaces mitosis specific  
 CC cohesin complex when it dissociates from chromatin during prophase  
 CC I. (By similarity).  
 CC -1- SUBUNIT: Component of the meiosis-specific cohesin complex, which  
 CC also contains the SMC1 (SMC1L1 or SMC1L2) and SMC3 heterodimer.  
 CC Such complex likely contains RAD21, or the meiosis-specific  
 CC related protein REC8 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.  
 CC In prophase I stage of meiosis, it is found along the axial  
 CC elements of synaptonemal complexes. In late-pachytene-diplotene,  
 CC the bulk of protein dissociates from the chromosome arms probably  
 CC because of phosphorylation by PLK, except at centromeres, where  
 CC cohesin complexes remain. It however remains chromatin associated  
 CC at the centromeres up to metaphase I. During anaphase I, it  
 CC probably dissociates from centromeres, allowing chromosomes







```
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
KW ubl conjugation pathway; Ligase; Multigene family; Mitosis;
FT BINDING 114 114
FT MUTAGEN 114 114 C->S: INHIBITION OF CYCLIN B DEGRADATION.
FT STRAND 114 114 UBIQUITIN.
FT HELIX 31 45
FT TURN 48 49
FT STRAND 50 54
FT TURN 59 60
FT STRAND 61 67
FT TURN 70 71
FT STRAND 73 76
FT TURN 78 84
FT STRAND 87 90
FT TURN 95 98
FT STRAND 104 104
FT TURN 105 106
FT STRAND 107 107
FT TURN 109 110
FT STRAND 112 113
FT TURN 116 118
FT HELIX 119 121
FT TURN 124 125
FT STRAND 128 138
FT TURN 139 140
FT HELIX 150 155
FT TURN 156 157
FT HELIX 159 175
SQ SEQUENCE 177 AA; 20134 MW; 06A454E798EB4F3A CRC64;

Query Match
Best Local Similarity 5.5%; Score 122.5; DB 1; Length 177;
Matches 45; Conservative 35; Mismatches 70; Indels 33; Gaps 10;

QY 234 KKNQDYLVNGVSSVQATRLMKELRDYRS-----QSPKGGNYAVELVNDLSYDNV 288
Db 15 QKERPRDVTTSKERHSV--SKRLQOELRTLMLSGDPGITAIPDG-----DNLFKV-V 63
QY 289 KLLKVDQPSALHNDLQILKEGADFIILNFSKDNFPDPFVVRVSPVLSGGVVLGGG 348
Db 64 ATLDPKDTVTES-----LTKYK-----LTLEFSDYDKPPVVKFTPCWHPN-VDQSG 111
QY 349 AICMELLPKQGWSSAYSIESVIMQISATLVKRGKARVQFGANKSQYSLTRAQQSKYSLVQI 408
Db 112 NICLDIL-KENWTASYDVRTLLSLQSL--GEPNNSPLNAQAADMSNQTKEYKV--L 166
QY 409 HEK 411
Db 167 HEK 169

RESULT 8
AGOL_ARATH STANDARD; PRT; 1048 AA.
AC 004379;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute protein.
GN AGOL OR ATIG48410 OR FLIA17.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Leaf;
RX MEDLINE=98090460; PubMed=9427751;
RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
```

```
RT *AGOL defines a novel locus of Arabidopsis controlling leaf
RT development.*;
RL EMBO J. 17:170-180(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Convey A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
RA Militsher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC STAGES.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U91995; AAC18440.1; -
CC EMBL; AC007932; AAD49755.1; -
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS50821; PAZ; 1.
CC PROSITE; PS50822; PIWI; 1.
CC Developmental protein.
CC DOMAIN 391 501 PAZ.
CC DOMAIN 676 997 PIWI.
CC DOMAIN 13 104 GLY-RICH.
CC SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;

Query Match
Best Local Similarity 5.5%; Score 122.5; DB 1; Length 1048;
Matches 87; Conservative 46; Mismatches 137; Indels 171; Gaps 21;

QY 3 QPQPGQQQPGP-----GQQLGG-----QGAAPGAGGGGPGGPGPCLRLKLESIHFG 55
Db 66 QPQQQQQYGGPQYQGRGGPPHQGRGGYGGGGGSSGP----- 109
QY 56 HERFRIASACLDLSCFLLAGAGAGAGAPGPHLPGRGVPDVRHCHNITESPAN 115
Db 110 -----PQRQSV-----ELQAT 122
QY 116 PPIWSVESDDNLAALVRLVDIKGNITLLQHLKRI-----ISDLCKLYNL 162
Db 123 SPTYQAVSSQPTSEVSPQV---PEPTVLAQOFELSVQEGAPSOAIQIPSSSKAFKF 179
```

RESULT 9

UBC6\_MOUSE STANDARD; PRT; 193 AA.

UBC6\_MOUSE  
P52482;

OC 01-OCT-1996 (Rel. 34, Created)

DT DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Ubiqutin-conjugating enzyme E2 E1 (EC 6.3.2.19) (Ubiqutin-protein  
ligase E1) (Ubiqutin carrier protein E1) (UBCH6).

GN UBE2E1 OR UBCH61 OR UBCM3 OR UBCES.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=Brain;

RC MEDLINE=96162026; PubMed=8576256;

RX Matuschewski K., Hauser H.P., Treier M., Jentsch S.;

RT "Identification of a novel family of ubiquitin-conjugating enzymes  
with distinct amino-terminal extensions.";

RL J. Biol. Chem. 271:2789-2794(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast tumor;

RC MEDLINE=23386257; PubMed=12477932;

RA Strausberg R.I., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeslee R., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER  
CC PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND  
CC ABNORMAL PROTEINS.

CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +  
CC diphosphate + protein N-ubiquityllysine.

```

RC TISSUE=Brain;
RX MEDLINE=98403980; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
RN [4]
RP SEQUENCE FROM N.A.
RA Wallis J., Lloyd C., Hall R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Binds to RAB11 and RAB25.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF395731; AAL12940.1; -
DR EMBL; AB014565; BAA31640.1; -
DR EMBL; AE006463; AAK61232.1; -
DR EMBL; AL023881; CAB92745.1; -
DR EMBL; AL049542; CAC17519.1; -
DR EMBL; Z98882; CAC17523.1; -
DR PIR; T00367; T00367.
DR GO; GO:0017137; F:RAB Interactor activity; NAS.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR SMART; SM00054; EFh; 2.
KW Calcium-binding; Repeat.
FT CA_BIND 215 226 EF-HAND 1 (POTENTIAL).
FT CA_BIND 247 258 EF-HAND 2 (POTENTIAL).
FT SEQUENCE 756 AA; 82439 MW; 264CEC399F28AFB9 CRC64;
SQ
Query Match 5.4%; Score 120; DB 1; Length 756;
Best Local Similarity 23.8%; Pred. NO. 3.7;
Matches 63; Conservative 28; Mismatches 104; Indels 70; Gaps 13;
QY 4 PQPGSQDQPGCQQLGG-----QGAAPGAGGSP-----GGGPGPG 38
Db 50 PQGLDGPAPGAADGAGSAGPAPGLGGPDPGPSAPPPRPGQASDPAPGP 109
QY 39 FCLRELKLLSIFRHRERFRIASACLDLSCEFLLAGAGAGAGA-----APGHP 92
Db 110 PRSEAPLELPDLSFTWTEPECPCASCP- SAPRFLGSSSHRACEVDVFSFP-AP 167
QY 93 PRGSV-----PGDPVRIHCNTESYPAVP--PIWSVESDDPNLAALVRLYDIKKNL 145
Db 168 TAGELALBQGGSPGPQ-PSDLSQTHP-LPSEFVGSQE-DGPRLRAVFDALDGDGDFVRI 224
QY 146 LQHLKRIISDLCKLNLPHQPDVEMLDQPL-----PAQCQTED 184
Db 225 EDFIQ-----FATVYGAEQVXDLTKYLDPSGLGVISFEDFYOGITATNGDPGCGYGV 279
QY 185 VSEDEDE-EMPEDEFDLDHYEMKE 208
Db 280 ASAQDEELPACDFEDDFVTEANE 304

```

```

RESULT 11
GCFC_HUMAN STANDARD; PRT; 917 AA.
AC Q9Y5B6; Q9NTQ0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE GC-rich sequence DNA-binding factor homolog.
GN GCFC OR C21ORF66.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,
RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA Antonarakis S.E.;
RT "From PREDs and open reading frames to cDNA isolation: revisiting the
RT human chromosome 21 transcription map.";
RL Genomics 78:46-54(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Brain, Heart, Kidney, Lung, Muscle, Placenta, and Testis;
RA Chapot-Skovgaard F.M., Guipponi M., Lyle R., Antonarakis S.E.;
RT "Isolation and initial characterization of a putative human chromosome
RT 21 transcription factor.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 132-917 FROM N.A. (ISOFORM A).
RX MEDLINE=20237674; PubMed=10773462;
RA Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N.,
RA Minoshima S., Kudoh J., Yaspo M., Ramser J., Reinhardt R., Reimer C.,
RA Clancy K., Rynditch A., Gardiner K.;
RT "Criteria for gene identification and features of genome organization:
RT analysis of 6.5 Mb of DNA sequence from human chromosome 21.";
RL Gene 247:215-232(2000).
RN [4]
RP SEQUENCE OF 489-815 FROM N.A. (ISOFORM B).
RA Teramoto T., Thorgeirsson S.S.;
RT "Cloning of candidate of GC-rich sequence DNA-binding factor.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Possible transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=Q9Y5B6-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q9Y5B6-2; Sequence=VSP_004267;
CC Name=C;
CC IsoId=Q9Y5B6-3; Sequence=VSP_004263, VSP_004264;
CC Name=D;
CC IsoId=Q9Y5B6-4; Sequence=VSP_004265, VSP_004266;
CC Note=Major isoform;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: TO GCFTCF9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY033903; AAK68721.1; -
DR EMBL; AY033904; AAK68722.1; -
DR EMBL; AY033905; AAK68723.1; -
DR EMBL; AY033906; AAK68724.1; -
DR EMBL; AJ279080; CAC40813.1; -
DR EMBL; AF231920; AAF72944.1; -

```

DR EMBL; AF153208; AAD34617.1; ALT\_TERM.  
DR Genew; HGNC:13579; C21orf66.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0003677; F:DNA binding activity; NAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW Alternative splicing.  
FT VARSPLIC 462 469 VPLINELE -> SVQFRKLL (in isoform C).  
FT VARSPLIC 470 917 /FTid-VSP\_004263.  
FT VARSPLIC 503 510 /FTid-VSP\_004264.  
FT VARSPLIC 511 917 /FTid-VSP\_004265.  
FT VARSPLIC 779 917 /FTid-VSP\_004266.  
FT VARSPLIC 917 917 /FTid-VSP\_004267.  
FT CONFLICT 692 711 MISSING (IN REF. 4).  
FT SEQUENCE 917 AA; 104803 MW; 855960A1D50A7789 CRC64;  
  
Query Match 5.4%; Score 120; DB 1; Length 917;  
Best Local Similarity 23.3%; Pred. No. 4.7;  
Matches 80; Conservative 38; Mismatches 115; Indels 110; Gaps 18;  
  
Qy 9 QQQPGGQQLGGGAAPGAGG---PGG---GPGPGPCLRLKLLLESIFHRGHERFRI 61  
Db 26 EQEPPPLPPGTGEEAGPGGDRAPGESLPGPSP----- 63  
Qy 62 ASACLDSECFELLAGAGGAGAGPPLPPRSGPDVRIHCNITESYPVPPPIWSV 121  
Db 64 PSALTPLGAB---AGGFPGGAEPGNGLKPR-KRPRE-----NKEVPRASLLSFQ 110  
Qy 122 ESDDPNLAALVRLVDIKGNTLLQLHLKRIISDLCKLYNLPHQPDVEM----- 170  
Db 111 DEEEN-----EEVFKVKKSS-----YSKIVKLLKREYK-----EDLEKSKIKTELNSA 156  
Qy 171 -LDQPLP-----AEQCTQED--VSSEDEDEEMPEDELDHYEMKEEPAEGKKSSEDGIG 223  
Db 157 ESEQPLDKTHVKDTNQEDGVIISEHGEDMDMESE-----KEEE-----KPKTGGA 204  
Qy 224 KENLAILEKIKKRODYLVNAGSVGSQVATDLKMLKELDIYRSQFKG-GNYAYELVND 282  
Db 205 SNALSSNLVRPGEIPD---AFTIAKKQKQARELGDFTPHDPNFGKGLVREDEDA 261  
Qy 283 LYDNVKKLLKVDPSALHNDLQILKEK-----EGAD 313  
Db 262 SDD-----EDDEKRRIVSVKESQKQKIAEIGIEGSD 296

RESULT 12  
ID UBC6\_HUMAN STANDARD; PRT; 193 AA.  
AC P51965;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Ubiquitin-conjugating enzyme E2 E1 (EC 6.3.2.19) (Ubiquitin-protein  
DE ligase E1) (Ubiquitin carrier protein E1) (Ubch6).  
CN UBC2E1 OR UBC6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96162027; Pubmed=8576257;  
RA Nuber U., Schwarz S., Kaiser P., Schneider R., Scheffner M.;  
RT "Cloning of human ubiquitin-conjugating enzymes Ubch6 and Ubch7  
(E2-F1) and characterization of their interaction with E6-AP and

RT RSP5";  
RL J. Biol. Chem. 271:2795-2800(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Colon;  
RX MEDLINE=22388257; Pubmed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., Morwan P.C., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -I- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER  
CC PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND  
CC ABNORMAL PROTEINS.  
CC -I- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +  
CC diphosphate + protein N-ubiquityllysine.  
CC -I- PATHWAY: Ubiquitin conjugation; second step.  
CC -I- SUBUNIT: INTERACTS WITH RNFI4.  
CC -I- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
CC thiolester formation.  
CC -I- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.  
CC -I- STRONGEST, TO DROSOPHILA UBCD2  
CC -----  
CC this SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X92963; CAA63539.1; -  
CC EMBL; BC009139; AAH09139.1; -  
CC HSP; P15731; IQCQ.  
CC Genew; HGNC:12477; UBE2E1.  
CC MIM; 602316; -  
CC GO; GO:0004840; F:ubiquitin conjugating enzyme activity; TAS.  
CC GO; GO:0004842; F:ubiquitin-protein ligase activity; TAS.  
CC InterPro; IPR000608; UBQ\_conjugat.  
CC Pfam; PF00179; UBQ\_con; 1-  
CC ProDom; PD000461; UBQ\_conjugat; 1.  
CC SMART; SM00212; UBCC; 1.  
CC PROSITE; PS00183; UBIQUITIN\_CONJUGAT\_1; 1.  
CC PROSITE; PS0127; UBIQUITIN\_CONJUGAT\_2; 1.  
KW Ubiquitin-conjugating pathway; Ligase; Multigene family.  
FT DOMAIN 9 18 POLY-SER.  
FT BINDING 131 131 UBIQUITIN (BY SIMILARITY).  
FT SEQUENCE 193 AA; 21404 MW; 2FBC50BE2A6A0008 CRC64;  
  
Query Match 5.4%; Score 119.5; DB 1; Length 193;  
Best Local Similarity 22.9%; Pred. No. 0.64;  
Matches 49; Conservative 32; Mismatches 96; Indels 37; Gaps 8;  
  
Qy 187 SEDEDEEMPEDELDHYEMKEEPAEGKKSDDGIGKENLAILEKIKKRODYLVNAGV 246  
Db 2 SDDDSRASTSSSSSSSSNOQTETPTPKKESKVSMSKSKLLTSKAK----- 50  
Qy 247 SGVQVATDRMLKMLKELDIYRSQFKGNYAYELVNDLSYDNVKKLLKVDPSALHNDLQIL 306

Db 51 -----RIKELADITLDP--PNC-SAGKGNINYEWRSTIL--GPPGSVY----- 91  
 QY 307 KEKAGADFTLLNFSKDPNPPPPRVVSPVLSGGYVGGGATCMEILLTQGNSSAYSII 366  
 Db 92 ---EGVFP-FLDITTPPPPKPKVTRTRIHCN-INSQVVICLDIL-KDNWSPALTI 145  
 QY 367 ESMVMOISATLVK-OKARVQFGANKSQYSLTRAQ 399  
 Db 146 SKVLLSICSLTDCNPADPLVGSATQYMTNRAE 179

RESULT 13  
 BAT8\_HUMAN  
 ID BAT8\_HUMAN STANDARD; PRT: 1210 AA.  
 AC Q96KQ7; Q14349; Q96MH5; Q96QD0; Q9UQL8; Q9Y331;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 3  
 DE (EC 2.1.1.43) (Histone H3-K9 methyltransferase 3) (H3-K9-HMTase 3)  
 DE (HLA-B associated transcript 8) (G9a) (NG36).  
 GN BAT8 OR G9A OR NG36.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L., Qin S., Madan A., Dickhoff R., Dors M., Madan A., Hicks P.,  
 RA Loretz C., Ratcliffe A., Abbasi N., Shaffer T., Hood L.;  
 RT "Sequence of the human major histocompatibility complex class III  
 region";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hirakawa M., Yamaguchi H., Imai K., Shimada J., Shina S., Tamiya G.,  
 RA Oka A., Inoko H.;  
 RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING (ISOFORM 2), AND  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=213564388; PubMed=11707778;  
 RA Brown S.E., Campbell R.D., Sanderson C.M.;  
 RT "Novel NG36/G9a gene products encoded within the human and mouse MHC  
 class III regions";  
 RL Mamm. Genome 12:916-924(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Salivary gland;  
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,  
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,  
 RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEO human cDNA sequencing project";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 5-1210 FROM N.A. (ISOFORM 1).  
 RC TISSUE=Muscle, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins E.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield A.S., Krzywinski M.I., Jones S.J.M., Marra M.A.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 195-1210 FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Histocytic lymphoma;  
 RX MEDLINE=93207535; PubMed=8457211;  
 RA Milner C.M., Campbell R.D.;  
 RT "The G9a gene in the human major histocompatibility complex encodes a  
 novel protein containing ankyrin-like repeats";  
 RL Biochem. J. 290:811-818(1993).  
 RN [7]  
 RP ENZYMIC ACTIVITY, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=21326082; PubMed=11316813;  
 RA Tachibana M., Sugimoto K., Fukushima T., Shinkai Y.;  
 RT "Set domain-containing protein, G9a, is a novel lysine-preferring  
 mammalian histone methyltransferase with hyperactivity and specific  
 selectivity to lysines 9 and 27 of histone H3";  
 RL J. Biol. Chem. 276:25309-25317(2001).  
 RN [8]  
 RP IDENTIFICATION IN COMPLEX WITH E2F6; TFDPI; MAX; MGA; EHM7ASE1; CBX3;  
 RP RING1; RNF2; MBLR; L3MBTL2 AND YAP2.  
 RX MEDLINE=21999559; PubMed=12004135;  
 RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;  
 RT "A complex with chromatin modifiers that occupies E2F- and  
 MYC-responsive genes in G0 cells";  
 RL Science 296:1132-1136(2002).  
 CC -|- FUNCTION: Histone methyltransferase. Preferentially methylates  
 Lys-9 of histone H3 and Lys-27 of histone H3 (in vitro). H3 Lys-9  
 methylation represents a specific tag for epigenetic  
 transcriptional repression by recruiting H3 proteins to  
 methylated histones. Probably targeted to histone H3 by different  
 DNA-binding proteins like E2F6, MGA, MAX and/or DPL. Also  
 methylates histone H1 (By similarity).  
 CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =  
 S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.  
 CC -|- SUBUNIT: Part of the E2F6-com-1 complex in G0 phase composed of  
 E2F6, MGA, MAX, TFDPI, CBX3, BAT8, EHM7ASE1, RING1, RNF2, MBLR,  
 L3MBTL2 and YAP2.  
 CC -|- SUBCELLULAR LOCATION: Nuclear; associates with euchromatic  
 regions. Does not associate with heterochromatin.  
 CC -|- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing: Named isoforms=3;  
 CC Comment-Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q96KQ7-1; Sequence=Displayed;  
 CC Name=2; Synonyms=NG36G9a-SPI;  
 CC IsoId=Q96KQ7-2; Sequence=VSP\_0022111;  
 CC Name=3; Synonyms=NG36;  
 CC IsoId=Q96KQ7-3; Sequence=VSP\_0022112, VSP\_0022113;  
 CC -|- TISSUE SPECIFICITY: Expressed in all tissues examined, with high  
 levels in fetal liver, thymus, lymph node, spleen and peripheral  
 blood leukocytes and lower level in bone marrow.  
 CC -|- SIMILARITY: BELONGS TO THE HISTONE-LYSINE METHYLTRANSFERASE  
 FAMILY. SUBFAMILY 3 SUBFAMILY.  
 CC -|- SIMILARITY: Contains 7 ANK repeats.  
 CC -|- SIMILARITY: Contains 1 pre-SET domain.  
 CC -|- SIMILARITY: Contains 1 SET domain.  
 CC -|- SIMILARITY: Contains 1 post-SET domain.  
 CC -|- CAUTION: NG36 and G9a were originally thought to derive from two  
 separate genes. Ref.3 shows that all G9a transcripts also contain  
 the in frame coding sequence of NG36.  
 CC -|- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to  
 erroneous gene model prediction.  
 CC -|- CAUTION: It is not known whether Met-1 or Met-21 is the initiator

methionine.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF134726; AAD21811.1; ALT\_SEQ.  
 EMBL: AF134726; AAD21812.1; ALT\_SEQ.  
 EMBL: AP000502; BAB63294.1; ALT\_SEQ.  
 EMBL: AP000502; BAB63295.1; ALT\_SEQ.  
 EMBL: AJ315532; CAC86666.1;  
 EMBL: AK056936; BAB71314.1;  
 EMBL: BC002686; AAH02686.1; ALT\_INIT.  
 EMBL: BC018718; AAH18718.1; ALT\_INIT.  
 EMBL: BC020970; AAH20970.1; ALT\_INIT.  
 EMBL: X69838; CAA49491.1; ALT\_INIT.  
 Genew; HGNC:14129; BAT8.  
 MIM; 604599;  
 InterPro; IPR002110; ANK.  
 InterPro; IPR001214; SET.  
 InterPro; IPR003606; Zn2-binding.  
 Pfam; PF00023; ank; 6.  
 Pfam; PF00856; SET; 1.  
 PRINTS; PR01415; ANKYRIN.  
 SMART; SM00468; PRESET; 1.  
 SMART; SM00317; SET; 1.  
 PROSITE; PS50088; ANK\_REPEAT; 5.  
 PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 PROSITE; PS50868; POST\_SET; FALSE\_NEG.  
 PROSITE; PS50867; PRE\_SET; 1.  
 PROSITE; PS50280; SET; 1.  
 Transferrase; Methyltransferase; Chromatin regulator; Nuclear protein;  
 ANK repeat; Repeat; Alternative splicing.  
 DOMAIN 2 13 POLY-ALA.  
 DOMAIN 160 163 POLY-ALA.  
 DOMAIN 300 326 POLY-GLU.  
 REPEAT 649 678 ANK 1.  
 REPEAT 684 713 ANK 2.  
 REPEAT 717 746 ANK 3.  
 REPEAT 750 780 ANK 4.  
 REPEAT 784 813 ANK 5.  
 REPEAT 817 846 ANK 6.  
 REPEAT 850 879 ANK 7.  
 DOMAIN 972 1035 PRE-SET.  
 DOMAIN 1037 1159 SET.  
 DOMAIN 1164 1180 POST-SET.  
 VARSPLIC 373 406 Missing (in isoform 2).  
 VARSPLIC 195 202 PPVEKRP -> VSGMGEMG (in isoform 3).  
 VARSPLIC 203 1210 /Frid-VSP\_002212.  
 VARSPLIC 5 5 Missing (in isoform 3).  
 CONFLICT 13 13 A -> Q (IN REF. 5; AAH18718).  
 CONFLICT 55 55 A -> G (IN REF. 5; AAH20970).  
 CONFLICT 168 169 N -> T (IN REF. 1 AND 4).  
 CONFLICT 178 178 PP -> TR (IN REF. 2; AAH02686).  
 CONFLICT 178 178 P -> S (IN REF. 3).  
 CONFLICT 985 985 C -> R (IN REF. 3).  
 CONFLICT 994 994 C -> R (IN REF. 6).  
 SEQUENCE 1210 AA; 132383 MW; E8508923CA66911A CRC64;  
 Query Match 5.38; Score 118.5; DB 1; Length 1210;  
 Best Local Similarity 23.4%; Pred. No. 8.5;  
 Matches 73; Conservative 34; Mismatches 120; Indels 85; Gaps 14;  
 QY 1 MOOPQOQOQPGGQ-----OLGGOGAAPGAGGPGGG-----PGPGCLRLRLKLL 48  
 Db 51 LPRANPDLSPAGSPSPSVTVVGDEGADTPVCAATPLIGDESNLEGGD-LRGGRILL 109

QY 49 ESTFHRGHERFRITASACLDLSC-----EFLLAGAG-----GAGACAAP 87  
 Db 110 -----GHATKSPSPSGKGGCPSPRAKMSMTGAGKSPSPSVQSLAMRLLSMPGAQAAAA 163  
 QY 88 GPHLPGRGVPDGVRIH-CNITESYP--AVPPI-----WSVESDDPNLAALVLER 134  
 Db 164 GSEPPATTSPEGQPVHRAKRTMSKPGNGQPPVPEKRPETIOHFRMSDDVHSLGKVTSD 223  
 QY 135 L-----VDIKKNTLLLOHLKRIISDLCKLYLNLPQHPDVM 170  
 Db 224 LAKRRKLSGGGLSEELSGARRSGEVTLTKDGPGLSEEWTVVGGDFSLVYDSYVD-ER 282  
 QY 171 LDQPLPAE-QCTQEDVSSDEDEMPEDTDLHYEMKEEPAE-CKKSEDDGIGKENLA 228  
 Db 283 VDSDSKSEVEALTEQLSEEEEEEEEEEEEEEEEEEEEEDEESGNSQDRSGSGR--- 339  
 QY 229 ILEKIKKNOROD 240  
 Db 340 --RKAKKKWRKD 349

RESULT 14  
 CIBX\_DROME  
 ID CIBX\_DROME STANDARD; PRT; 985 AA.  
 AC P17970; O76805; Q9I7T9; Q9I7U0;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel protein Shab.  
 GN SHAB OR CG1066.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN-Oregon-R;  
 RX MEDLINE=89146139; PubMed=24933160;  
 RA Butler A., Wei A.G., Baker K., Salkoff L.;  
 RT "A family of putative potassium channel genes in Drosophila.";  
 RL Science 243:943-947(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG).  
 RX MEDLINE=90239553; PubMed=2333511;  
 RA Wei A.G., Covarrubias M., Butler A., Baker K., Salkoff L.;  
 RT "K+ current diversity is produced by an extended gene family  
 conserved in Drosophila and mouse.";  
 RL Science 248:599-603(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG).  
 RX MEDLINE=90245668; PubMed=2336395;  
 RA Butler A., Wei A.G., Salkoff L.;  
 RT "Shal, Shab, and Shaw: three genes encoding potassium channels in  
 Drosophila.";  
 RL Nucleic Acids Res. 18:2173-2174(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND MUTAGENESIS.  
 RC STRAIN-Canton-S;  
 RX MEDLINE=99348337; PubMed=10419540;  
 RA Hegde P., Gu G.G., Chen D., Free S.J., Singh S.;  
 RT "Mutational analysis of the Shab-encoded delayed rectifier K(+) channels in Drosophila.";  
 RL J. Biol. Chem. 274:22109-22113(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RC STRAIN-Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abille J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellista A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM  
 CC ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED  
 CC CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE  
 CC MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH  
 CC WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL  
 CC GRADIENT.  
 CC -!- SUBUNIT: HETEROTRIMER OF POTASSIUM CHANNEL PROTEINS (PROBABLE).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms-2;  
 CC Name-Long;  
 CC IsoId=P17970-1; Sequences-Displayed;  
 CC Name-Short;  
 CC IsoId=P17970-2; Sequences-VSP\_000960;  
 CC -!- DEVELOPMENTAL STAGE: Expressed in late embryos and pupae.  
 CC -!- DOMAIN: THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE  
 CC RATE OF INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE  
 CC IN MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL  
 CC TO SPECIFIC SUBCELLULAR COMPARTMENTS.  
 CC -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND  
 CC IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
 CC EVERY THIRD POSITION.  
 CC -!- SIMILARITY: BELONGS TO THE POTASSIUM CHANNEL FAMILY. SHAB  
 CC SUBFAMILY.  
 CC -!- CAUTION: Ref.5 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; M32659; AAA2896.1; -;  
 CC DR EMBL; AF084525; AAC33365.1; -;  
 CC DR EMBL; AE003476; AAG22232.1; ALT\_SEQ.  
 CC DR EMBL; AE003476; AAG22233.1; ALT\_INIT.

DR PIR; S12746; S12746.  
 DR HSSP; Q54397; 1BL8.  
 DR FlyBase; FBgn0003383; Shab.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K+channel.  
 DR InterPro; IPR003131; Ktetra.  
 DR InterPro; IPR003971; Kv9\_channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR003820; M+channel\_nlg.  
 DR Pfam; PF00520; Ion\_trans; 1.  
 DR Pfam; PF02214; Ktetra; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01494; KV9CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium transport; Potassium; transmembrane;  
 KW Glycoprotein; Phosphorylation; Multigene family; Alternative splicing.  
 FT TRANSMEM 436 454 SEGMENT S1.  
 FT TRANSMEM 474 495 SEGMENT S2.  
 FT TRANSMEM 506 527 SEGMENT S3.  
 FT TRANSMEM 536 561 SEGMENT S4.  
 FT TRANSMEM 577 598 SEGMENT S5.  
 FT TRANSMEM 638 659 SEGMENT S6.  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 749 749 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 756 756 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 888 888 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT MOD\_RES 690 690 PHOSPHORYLATION (BY PKA) (POTENTIAL).  
 FT MOD\_RES 731 731 PHOSPHORYLATION (BY PKA) (POTENTIAL).  
 FT MOD\_RES 796 796 PHOSPHORYLATION (BY PKA) (POTENTIAL).  
 FT VARSPIC 717 746 Missing (in isoform Short).  
 FT /FTid=VSP\_000960.  
 FT R>O: IN ALLELE SHAB-1; TEMPERATURE-  
 FT SENSITIVE PARALYTIC.  
 FT V>D: IN ALLELE SHAB-1; TEMPERATURE-  
 FT SENSITIVE PARALYTIC.  
 FT Q -> L (IN REF. 1, 2 AND 3).  
 FT S -> G (IN REF. 1, 2 AND 3).  
 FT R -> G (IN REF. 1, 2 AND 3).  
 FT S -> G (IN REF. 1, 2 AND 3).  
 FT ES -> CA (IN REF. 5).  
 FT A -> S (IN REF. 1, 2 AND 3).  
 FT V -> I (IN REF. 5).  
 FT A -> T (IN REF. 5).  
 FT C -> Y (IN REF. 5).  
 FT V -> I (IN REF. 5).  
 FT EQ -> DE (IN REF. 1, 2 AND 3).  
 FT A -> G (IN REF. 1, 2 AND 3).  
 FT GDGGGGVDDNLSOAKGLPIOMITGVEAEALRRQVALEN  
 FT LQNQMDNLEQDVFVECCCTTKGLPGHCICPLRANS  
 FT V -> VMEMGAVSMTTTPRPRDCPSR (IN REF. 1, 2  
 FT AND 3).  
 FT DDD -> NDN (IN REF. 4).  
 FT SEQUENCE 911 913 106359 MW; 59E38AD35F064AC8 CRC64;  
 SQ  
 Query Match 5.3%; Score 118; DB 1; Length 985;  
 Best Local Similarity 23.5%; Pred. No. 6.9;  
 Matches 63; Conservative 27; Mismatches 102; Indels 76; Gaps 12;  
 QY 13 GPGQOLGGQGAAPGAGGGPGGPGP-CLRELKLLSFHGRHFRIASACLD-ELS 70  
 Db 163 GAGASVTGSGGAGTGTGTGAGSGSGGAGAEKVR-----YAPFPVASPTHSIPTT 213  
 QY 71 CEFLLAGAGGAGAGAAAGPHLPGRGVP-----GDPVRIHCNITESYPAVPPTNSVE 122  
 Db 214 SQQIVGSGVGVGVGGGASSQSI--SGGVPTHSQSNSTGALQRTHSRSMSSIPPPFFMTAQ 271



QY	408	IHEK	411
		:	:
Db	166	LHEO	169

Search completed: July 29, 2003, 13:45:22  
Job time : 25 secs

[illegible]



Db 28 SFHIELEDDSNFTWNGVNLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76  
Qy 332 VRVSPVLGGYVGGGAGCMLTKOG-----WSSAYSIESVIMQISATL--- 377  
Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPQTVESVLISVSLLEDP 135  
Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410  
Db 136 NINSPANVDAVDYRKNPQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195  
Qy 411 -----KNGWYTPPKED 421  
Db 196 NKDMADNEWYDSLDLD 211

RESULT 2  
US-09-196-525-5  
; Sequence 5, Application US/09196525  
; Patent No. 5989883  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09196,525  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/679,765  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0093 US  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 788905  
US-09-196-525-5

Query Match 6.1%; Score 135.5; DB 2; Length 295;  
Best Local Similarity 21.4%; Pred. No. 0.0012;  
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;

Qy 273 NYAVELVNDLS-LYDNNVLLKVDODSALHNDLQILKEGADFIILLNFSKDNRPDPFP 331  
Db 28 SFHIELEDDSNFTWNGVNLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76  
Qy 332 VRVSPVLGGYVGGGAGCMLTKOG-----WSSAYSIESVIMQISATL--- 377

Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPQTVESVLISVSLLEDP 135  
Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410  
Db 136 NINSPANVDAVDYRKNPQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195  
Qy 411 -----KNGWYTPPKED 421  
Db 196 NKDMADNEWYDSLDLD 211

RESULT 3  
US-09-318-317-5  
; Sequence 5, Application US/09318317  
; Patent No. 6172199  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/318,317  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/196,525  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0093 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 788905  
US-09-318-317-5

Query Match 6.1%; Score 135.5; DB 3; Length 295;  
Best Local Similarity 21.4%; Pred. No. 0.0012;  
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;

Qy 273 NYAVELVNDLS-LYDNNVLLKVDODSALHNDLQILKEGADFIILLNFSKDNRPDPFP 331  
Db 28 SFHIELEDDSNFTWNGVNLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76  
Qy 332 VRVSPVLGGYVGGGAGCMLTKOG-----WSSAYSIESVIMQISATL--- 377  
Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPQTVESVLISVSLLEDP 135  
Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410

Db 136 NINSPANVDAADYRKNEQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195  
QY 411 -----KNGWYTPPKED 421  
Db 196 NKDMADNFWYDSLDLD 211

## RESULT 4

US-09-177-165A-22  
; Sequence 22, Application US/09177165A  
; Patent No. 6426205  
; GENERAL INFORMATION:  
; APPLICANT: Tvers, Mike  
; APPLICANT: Willems, Andrew  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN  
; TITLE OF INVENTION: DEPENDENT PROTEOLYSIS  
; FILE REFERENCE: 11757.10USU1  
; CURRENT APPLICATION NUMBER: US/09/177.165A  
; CURRENT FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/092,443  
; PRIOR FILING DATE: 1998-07-10  
; PRIOR APPLICATION NUMBER: 60/063,254  
; PRIOR FILING DATE: 1997-10-24  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-177-165A-22

Query Match 6.1%; Score 135.5; DB 4; Length 295;  
Best Local Similarity 21.4%; Pred. No. 0.0012;  
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;  
QY 273 NYAVELVND-SLYDNVKKLVQDQDSALHNDLQILKEGADFILLNFSKDNFPDPFP 331  
Db 28 SFHIELEDDSNIFWNGVWLNEDSIH-----CGFFKAQMRFPDPFP 76  
QY 332 VRVVSPVLGGYVGGGAIKMLITKQ-----WSSAYSIESVIMQISATL--- 377  
Db 77 FR-FTPAIYHPNVYRDGRICISILHQSGDPMTDEPDAETWSPQTVESVLISIVSLLEDP 135  
QY 378 -----VRGKARVQFCAKNSQYS-----LTRAQOSYKSLVQIHE--- 410  
Db 136 NINSPANVDAADYRKNEQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195  
QY 411 -----KNGWYTPPKED 421  
Db 196 NKDMADNFWYDSLDLD 211

## RESULT 5

US-08-828-533-3  
; Sequence 3, Application US/08828533  
; Patent No. 6180379  
; GENERAL INFORMATION:  
; APPLICANT: Ruderman, Joan V.  
; APPLICANT: Hershko, Avram  
; APPLICANT: Kirschner, Marc W.  
; APPLICANT: Townsley, Fiona  
; APPLICANT: Aristarkov, Alexander  
; APPLICANT: Eytan, Esther  
; APPLICANT: Yu, Hongtao  
; TITLE OF INVENTION: NOVEL CYCLIN-SELECTIVE UBIQUITIN CARRIER  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: United States of America

ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/828,533  
FILING DATE: 31-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kerner Ph.D., Ann-Louise  
REGISTRATION NUMBER: 33,523  
REFERENCE/DOCKET NUMBER: HAZ-01SCIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 178 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: linear  
FRAGMENT TYPE: linear  
US-08-828-533-3

Query Match 5.5%; Score 122.5; DB 3; Length 178;  
Best Local Similarity 24.6%; Pred. No. 0.0076;  
Matches 45; Conservative 35; Mismatches 70; Indels 33; Gaps 10;  
QY 234 KKNQDYLNGAVSGSVQATDRLMKELRDIYS-----QSFKGGNAYVELVNDSDVNWY 288  
Db 15 QKERPRDMTTSKERHSV--SKRLQQLRLTLLMSGDPCITAFPDG-----DNLEFKV-V 63  
QY 289 KLLKVQDSDALHNDLQILKEGADFILLNFSKDNFPDPFPVVRVSPVLGGYVGGG 348  
Db 64 ATLDGPKDVTYES-----LKYK-----LTFLEPFDYIPKPPVVKVTTTCWHFN-V 111  
QY 349 AICMELITKQWSSAYSIESVIMQISATLVKGRVQFGANKSQYSILTRAQOSYKSLVQI 408  
Db 112 NICLDIL-KENWTASYDVTRILLSLQSL--GEPNNASPLNAQAADWWSNQTEYKKV--L 166  
QY 409 HEK 411  
Db 167 HEK 169

## RESULT 6

US-09-772-156-3  
; Sequence 3, Application US/09772156  
; Patent No. 6528633  
; GENERAL INFORMATION:  
; APPLICANT: Ruderman, Joan V.  
; APPLICANT: Hershko, Avram  
; APPLICANT: Kirschner, Marc W.  
; APPLICANT: Townsley, Fiona  
; APPLICANT: Aristarkov, Alexander  
; APPLICANT: Eytan, Esther  
; APPLICANT: Yu, Hongtao  
; TITLE OF INVENTION: NOVEL CYCLIN-SELECTIVE UBIQUITIN CARRIER  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: United States of America  
; ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/772,156
; FILING DATE: 31-Mar-1997
; CLASSIFICATION: A35
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,533
; FILING DATE: 1997-03-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner Ph.D., Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: HAL-015CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: <Unknown>
; FRAGMENT TYPE: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
; US-09-772-156-3

```

Query Match	5.5%;	Score 122.5;	DB 4;	Length 178;
Best Local Similarity	24.6%;	Pred. No. 0.0076;		
Matches	45;	Conservative 35;	Mismatches 70;	Indels 33; Gaps 10;
QY	234	KKNQDQYLNGAVSGSYQATDRLMKELRDIYRS-----QSFKGGYAVELYNDLSLYDNV	288	
Db	15	QKERPRDWTTSKERHSV--SKRQQLUETLIMGDPGIGTAFPDG-----DNLFRK-V	63	
QY	289	KLLKVDQDSALHNDLQILKEGADFILLNFSFKDNFPDPFVRVSPVLSGGYVLGGG	348	
Db	64	ATLDGPKDTVYES----LKYK-----LTLEPDSYPYKPPVWFPTPCWHPN-VDQSG	111	
QY	349	AICMELLTQKGSWSSAYSIESVIMQISATLYKGARYQFGANKSQVSLYRPAQSYKSLVQI	408	
Db	112	NICLDIL-KENWTASYDVRTILLSQLL--GEPNNASPLNQAADWMSNOTYEKKV--L	166	
QY	409	HEK 411		
Db	167	HEK 169		

```

RESULT 7
US-07-720-589-2
; Sequence 2, Application US/07720589
; Patent No. 5324630
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; APPLICANT: Perng, Guey-Chen
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/720,589
; FILING DATE: 19910628

```

[illegible]

```

RESULT 8
US-08-785-190-2
; Sequence 2, Application US/08785190
; Patent No. 5977339
;
GENERAL INFORMATION:
;
; APPLICANT: Lefebvre, Rance B.
; APPLICANT: Perng, Guey-Chen
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
;
CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
;
COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,190

```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/720,589
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-785-190-2

```

```

Query Match      5.4%; Score 120.5; DB 2; Length 700;
Best Local Similarity 21.3%; Pred. No. 0.085;
Matches 80; Conservative 62; Mismatches 133; Indels 101; Gaps 18;

QY 90 HLPGRGVPDPVRIHCNITESYPVPPVWSVSDPNLAALVRLVYDIKKGNLTLLQHL 149
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 339 NLPKPGDVSPKVDKQIQIKESLEDLQEQLK-ETGDENQKREIEKQIEIKKSDKLLK-- 395
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 KRIISDLCKLYNLQHPDVEMLDQPLPAEQCTQEDVSSDEP-----DEEMPEDTEDLDHY 204
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 -----SKDKASKDGKALDLDRELNSKASSKEKS 424
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 205 EMKEEPAEGK--KSEDDGIGKENLAI-----LEKIKK-NORQDYLNGAVSGSVQATDRL 256
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 425 KAKEEITTKGSKSLGDLNNDENLAMPEDQKLPVKKLSKKEF---KPVSEVEKLDKI 481
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 MK-----ELRDIYRSQSPKGGNYAVELVNDSDYVNVKLLKVDQDSALHNDLQILKE 310
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 FKSNNVNGELSPDKS-SYKDIDSKEETVKNQKTPQVQKDVTSLENDLTTMSIDS 540
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 GADPILLNFSEKDNFPDPFVRVSPVLSGGVY-----LGGGAICMELLTKQGWS--SAY 364
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 541 SS-----PVFLEVIDPTNLGLQLIDLNTG-VSLKESTQGGIORYGIY 583
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 SIESVIMQISATLVKGRVQF-----GANKSQYSLTRAQOQSYKS-----LVQIHEK- 411
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 584 EREKDLVVIK--MDSGKAKQLIDKLENLKVVSSENFINKNSSLYVDSKMLVAVRDKD 641
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 412 --NGW-----YTPPKED 421
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 SSNDWRLAKFSPKNLD 657

```

```

RESULT 9
PCT-US92-05539-2
; Sequence 2, Application PC/TUS9205539
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; APPLICANT: Perring, Guey-Chen
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05539
; FILING DATE: 19920629
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-05539-2

```

```

Query Match      5.4%; Score 120.5; DB 5; Length 700;
Best Local Similarity 21.3%; Pred. No. 0.085;
Matches 80; Conservative 62; Mismatches 133; Indels 101; Gaps 18;

QY 90 HLPGRGVPDPVRIHCNITESYPVPPVWSVSDPNLAALVRLVYDIKKGNLTLLQHL 149
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 339 NLPKPGDVSPKVDKQIQIKESLEDLQEQLK-ETGDENQKREIEKQIEIKKSDKLLK-- 395
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 KRIISDLCKLYNLQHPDVEMLDQPLPAEQCTQEDVSSDEP-----DEEMPEDTEDLDHY 204
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 -----SKDKASKDGKALDLDRELNSKASSKEKS 424
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 205 EMKEEPAEGK--KSEDDGIGKENLAI-----LEKIKK-NORQDYLNGAVSGSVQATDRL 256
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 425 KAKEEITTKGSKSLGDLNNDENLAMPEDQKLPVKKLSKKEF---KPVSEVEKLDKI 481
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 MK-----ELRDIYRSQSPKGGNYAVELVNDSDYVNVKLLKVDQDSALHNDLQILKE 310
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 FKSNNVNGELSPDKS-SYKDIDSKEETVKNQKTPQVQKDVTSLENDLTTMSIDS 540
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 GADPILLNFSEKDNFPDPFVRVSPVLSGGVY-----LGGGAICMELLTKQGWS--SAY 364
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 541 SS-----PVFLEVIDPTNLGLQLIDLNTG-VSLKESTQGGIORYGIY 583
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 SIESVIMQISATLVKGRVQF-----GANKSQYSLTRAQOQSYKS-----LVQIHEK- 411
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 584 EREKDLVVIK--MDSGKAKQLIDKLENLKVVSSENFINKNSSLYVDSKMLVAVRDKD 641
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 412 --NGW-----YTPPKED 421
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 SSNDWRLAKFSPKNLD 657

```

```

RESULT 10
US-08-679-765-3
; Sequence 3, Application US/08679765
; Patent No. 5840866
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```



```

, , REFERENCE/DOCKET NUMBER: PF-0093 US
, ,
, , TELECOMMUNICATION INFORMATION:
, ,
, , TELEPHONE: 415-855-0555
, ,
, , TELEFAX: 415-845-4166
, ,
, , INFORMATION FOR SEQ ID NO: 3:
, ,
, , SEQUENCE CHARACTERISTICS:
, ,
, , LENGTH: 193 amino acids
, ,
, , TYPE: amino acid
, ,
, , STRANDEDNESS: single
, ,
, , TOPOLOGY: linear
, ,
, , MOLECULE TYPE: peptide
, ,
, , IMMEDIATE SOURCE:
, ,
, , LIBRARY: GenBank
, ,
, , CLONE: 1064914
, ,
, , US-09-318-317-3

```

Query Match	5.4%	Score 119.5	DB 3	Length 193
Best Local Similarity	22.9%	Pred. No. 0.015		
Matches	49	Conservative	32	Mismatches 96
			Indels	37
			Gaps	8
QY	187	SEDEDEMPEDTEDLDHYEMKEEPEAGKKSSEDGICKENLAILEKIKKQRDYLN	GA	246
		: :	: :	: :
Db	2	SDDDSRASTSSSSSSNQOTKETNTPPKKESKVSMSKNKLLSTSAK		50
QY	247	SGSVQATDRMLKELDIYRSQSKGGNYAYELVNDSLYDNWKLLKVDQDSALHNDLQIL		306
		: :	: :	: :
Db	51	-----RIQKELADITLDP-----PNC\$AGPKGDNIEYK\$TIL-----GPF\$GVY-----		91
QY	307	KEKEGADFILLN\$F\$KDNPPDPFPYRVV\$PVL\$GGVYLVGGGAICMELLTKQGW\$SN\$Y\$		366
		: :	: :	: :
Db	92	---EGGVF-FLDITTFPEYFKPKPVTFTRIYHCN-INSQGVICLDIL-KDNW\$PALTI-		145
QY	367	ESVIMQISATLVK-GRARVOFGANK\$QY\$S\$TRAQ		399
		: :	: :	: :
Db	146	SKVLL\$IC\$SLTDCNPADPLV\$G\$ATQYMTNRAE		179

```

1  RESULT 13
2  US-08-679-765-1
3  ; Sequence 1, Application US/08679765
4  ; Patent No: 5840866
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Au-Young, Janice
7  ; APPLICANT: Goli, Surya K.
8  ; APPLICANT: Hillman, Jennifer L.
9  ; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING ENZYME
10 ; NUMBER OF SEQUENCES: 5
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
13 ; STREET: 3174 Porter Drive
14 ; CITY: Palo Alto
15 ; STATE: CA
16 ; COUNTRY: U.S.
17 ; ZIP: 94304
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: Diskette
20 ; COMPUTER: IBM Compatible
21 ; OPERATING SYSTEM: DOS
22 ; SOFTWARE: FastSeq Version 1.5
23 ; CURRENT APPLICATION DATA:
24 ; APPLICATION NUMBER: US/08/679,765
25 ; FILING DATE: Filed Herewith
26 ; ATTORNEY/AGENT INFORMATION:
27 ; NAME: Ballings, Lucy J.
28 ; REGISTRATION NUMBER: 36,749
29 ; REFERENCE/DOCKET NUMBER: PF-0093 US
30 ; TELECOMMUNICATION INFORMATION:
31 ; TELEPHONE: 415-855-0555
32 ; TELEFAX: 415-845-4166
33 ; INFORMATION FOR SEQ ID NO: 1:
34 ; SEQUENCE CHARACTERISTICS:
35 ; LENGTH: 201 amino acids
36 ; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: PGANNOT01
; CLONE: 61887
US-08-679-765-1

```

	Query Match		5.38;	Score 119;	DB 2:	Length 201;
	Best local Similarity	21.1%;	Pred.	No. 0.018;		
	Matches	45;	Conservative	42;	Mismatches	90;
		Indels	36;	Gaps	9;	
QY		189	DEDEMPDEDLDHYEMKEEPAEGKSKSDDDGIGKENLAILBKIKKNORODYLNGAVSG	248		
Db			:   :   :   :     :	: ~~~~~~	:	:  :
		9	DDSPTSGGSDGDRESVQQEPER-----EQVPKKKGKISKSTAA	51		
QY		249	SVO-ATDRLMKELRDYRSOFKGNAYVELVNDSLYDNWVKLLRVKDQDQSALHNDQLILK	307		
Db			: : :    :    :    :	:  :		
		52	KLSTSARKIQELAETIDLPDP---PNC SAGPKGDNIYEWRSTIL--GPPEGSVI-----	99		
QY		308	EKEGADFILLNFSEFKNFPDPFVRVSPVLSSGVYGLVGATCMELLRFQGHSASYSIE	367		
Db			:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	:  :  :  :  :  :  :		
		100	--EGGYF-FLDITFSPDYPFKPKVFTRIRYHCN-INSGVICLDLIL-KDNWSPALTTIS	154		
QY		368	SVIMQISATIV-RKGARVQFGANKSQSYSLTRAQ	399		
Db			:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	:  :  :  :  :  :  :		
		155	KVILLSICSLTDXNPADPLVGSSTATQYMTNRAE	187		

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```



